GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: January 30, 2004, 13:12:38 ; Search time 22 Seconds
(without alignments)
2313.635 Million cell updates/sec

Title: US-09-769-699-2
Perfect score: 6294
Sequence: 1 MENTQKTVTVPTGPLGYVYA.....DELFDLSGIPIKHGNITMEM 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3054	48.5	1452	4	US-09-127-227-2	Sequence 2, Appli
2	183	2.9	35	2	US-08-807-332B-9	Sequence 9, Appli
3	183	2.9	35	3	US-09-338-876-9	Sequence 9, Appli
4	131	2.1	757	3	US-09-413-814-84	Sequence 84, Appl
5	123	2.0	1074	4	US-09-071-035-358	Sequence 358, App
6	123	2.0	1074	4	US-09-071-035-394	Sequence 394, App
7	118.5	1.9	370	4	US-09-252-991A-16913	Sequence 16913, A
8	111	1.8	5215	3	US-09-105-537-2	Sequence 2, Appli
9	106.5	1.7	3567	2	US-07-642-734C-4	Sequence 4, Appli
10	106.5	1.7	3567	3	US-08-439-009A-4	Sequence 4, Appli
11	106	1.7	774	1	US-07-731-157A-7	Sequence 7, Appli
12	106	1.7	774	2	US-08-541-780-7	Sequence 7, Appli
13	105.5	1.7	1095	4	US-09-107-532A-3855	Sequence 3855, Ap
14	105	1.7	774	1	US-08-633-760-46	Sequence 46, Appl
15	105	1.7	774	1	US-08-633-760-48	Sequence 48, Appl
16	104.5	1.7	635	4	US-08-931-608A-5	Sequence 5, Appli
17	104.5	1.7	774	1	US-08-019-870-5	Sequence 5, Appli
18	103	1.6	774	1	US-08-019-870-3	Sequence 3, Appli
19	103	1.6	774	1	US-08-633-760-44	Sequence 44, Appl
20	102	1.6	729	1	US-08-070-165F-6	Sequence 6, Appli
21	102	1.6	729	2	US-08-885-418-6	Sequence 6, Appli
22	102	1.6	774	1	US-07-747-901A-3	Sequence 3, Appli
23	102	1.6	774	1	US-07-935-312-3	Sequence 3, Appli
24	102	1.6	774	1	US-08-633-760-50	Sequence 50, Appl
25	101.5	1.6	773	1	US-08-019-870-1	Sequence 1, Appli
26	101.5	1.6	773	1	US-08-019-870-6	Sequence 6, Appli
27	101.5	1.6	774	1	US-08-019-870-8	Sequence 8, Appli
28	101.5	1.6	774	1	US-08-019-870-11	Sequence 11, Appl
29	101.5	1.6	774	1	US-08-633-760-52	Sequence 52, Appl
30	101	1.6	995	4	US-09-657-931A-1	Sequence 1, Appli
31	101	1.6	1024	4	US-09-562-737-48	Sequence 48, Appl
32	101	1.6	1466	4	US-09-262-537-20	Sequence 20, Appl
33	101	1.6	7257	3	US-09-335-409-5	Sequence 5, Appli
34	101	1.6	7257	4	US-09-568-102-5	Sequence 5, Appli
35	101	1.6	7257	4	US-09-567-969-5	Sequence 5, Appli
36	101	1.6	7257	4	US-09-568-480-5	Sequence 5, Appli
37	101	1.6	7257	4	US-09-568-486-5	Sequence 5, Appli
38	101	1.6	7257	4	US-09-568-472-5	Sequence 5, Appli
39	101	1.6	7257	4	US-09-567-899-5	Sequence 5, Appli
40	100.5	1.6	1321	1	US-08-261-822A-3	Sequence 3, Appli
41	100.5	1.6	1321	5	PCT-US95-07744A-3	Sequence 3, Appli
42	100.5	1.6	2616	6	5206163-3	Patent No. 5206163
43	99.5	1.6	600	4	US-09-252-991A-29817	Sequence 29817, A
44	99	1.6	915	1	US-08-328-322-5	Sequence 5, Appli
45	98.5	1.6	801	4	US-09-351-150A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-127-227-2
; Sequence 2, Application US/09127227
; Patent No. 6399354
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; TITLE OF INVENTION: Fusion Protein
; FILE REFERENCE: HU98-05
; CURRENT APPLICATION NUMBER: US/09/127,227
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-09-127-227-2

Query Match 48.5%; Score 3054; DB 4; Length 1452;
Best Local Similarity 49.8%; Pred. No. 2.7e-304;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

```
Qy          1 MENTQKTVT---VPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVE 55
              ||   || |   || |||||||  ||  | ::|  :: |:||| |:|:|: ||:  ||||
Db          1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSGDADVAVAPLVVGLTVE 58

Qy         56 KTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNA 115
                | :::||| |:||||| |  ::|||| ||:  ||:|||||:|: ||: ||||||  |
Db         59 SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCER 118

Qy        116 ARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175
              ||  ||||      | |   |||| :| ||||:|:  :|||||  ||||| : | |||
Db        119 ARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINNTFLHL 178

Qy        176 GGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCH 235
              || | | |  :| |||::|:|||||| :|:: :||| :|||||| | || ||:|  |
Db        179 GGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFNRPLNR 238

Qy        236 LIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTAR 295
              |: : |: | ||||| ||| |||| |||||||||||||| || |||:| |:::|     | |
Db        239 LLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KTPR 296

Qy        296 GARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355
              | |           |:||||:||||:|| | || | | :  :::| ||||  ||:  |
Db        297 GGR-----DGGGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLCEGQ 351

Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415
```

09/769,699
2-21-04
SEARCH NOTES

```
Db        352 DTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLV 410

Qy        416 AGPHLAANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYL 468

Db        411 PGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYL 469

Qy        469 ERCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQAT 527

Db        470 ERCDGGVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAA 529

Qy        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587

Db        530 RGAIGVFGTMNSMYSDCDVLGNYAAFSALKR-ADGSETARTIMQETYRAATERVMAELET 588

Qy        588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647

Db        589 LQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLG 648

Qy        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707

Db        649 EANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQ 708

Qy        708 PVLRRRFVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766

Db        709 PVLRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFP 768

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826

Db        769 KELRVKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFP 828

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886

Db        829 NGKPPGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFINLAPNNV 888

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946

Db        889 SELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAW---SAQGGAGLEAGA 945

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006

Db        946 RALMDAVDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWAS 1005

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066

Db       1006 LMGGKNACPLLIFDRTRKFVLACPRAGFVCAASNLGGGAHESSLCEQLRGIISEGGAAVA 1065

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL--- 1123

Db       1066 SSVFVATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVA 1125

Qy       1124 -EAVKILDEKTTAGDGETPTNLAFNFD--SCEPSHDTTSNVLNISGSNISGSTVPGLKRP 1180

Db       1126 HEAEALVSQLGNAGE-------VFNFGDFGCE-----DDNATPFGGPGAPGPAFAGRKRA 1173

Qy       1181 PEDDELFDLSGIPIKHGNITMEM 1203

Db       1174 FHGDDPFG-EGPPDKKGDLTLDM 1195


RESULT 2
US-08-807-332B-9
; Sequence 9, Application US/08807332B
; Patent No. 5959074
;  GENERAL INFORMATION:
;    APPLICANT:  Dreyfus, David H.
;    APPLICANT:  Gelfand, Erwin W.
;    TITLE OF INVENTION:  PRODUCTS AND PROCESSES FOR REGULATION OF
;    TITLE OF INVENTION:  GENE RECOMBINATION
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross
;      STREET:  1700 Lincoln St., Suite 3500
;      CITY:  Denver
;      STATE:  CO
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/807,332B
;      FILING DATE:  28-FEB-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kovarik, Joseph E.
;      REGISTRATION NUMBER:  33,005
;      REFERENCE/DOCKET NUMBER:  2879-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  303/863-9700
;      TELEFAX:  303/863-0223
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-807-332B-9

  Query Match             2.9%;  Score 183;  DB 2;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-11;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        545 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 579
              |||||||||||||||||||||||||||||||||||
Db          1 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 35


RESULT 3
US-09-338-876-9
; Sequence 9, Application US/09338876
; Patent No. 6187584
;  GENERAL INFORMATION:
;    APPLICANT:  Dreyfus, David H.
;    APPLICANT:  Gelfand, Erwin W.
;    TITLE OF INVENTION:  PRODUCTS AND PROCESSES FOR REGULATION OF
;    TITLE OF INVENTION:  GENE RECOMBINATION
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross
;      STREET:  1700 Lincoln St., Suite 3500
;      CITY:  Denver
;      STATE:  CO
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/338,876
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/807,332
;      FILING DATE:  28-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kovarik, Joseph E.
;      REGISTRATION NUMBER:  33,005
;      REFERENCE/DOCKET NUMBER:  2879-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  303/863-9700
```

```
      TELEFAX:  303/863-0223
  INFORMATION FOR SEQ ID NO:  9:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  35 amino acids
      TYPE:  amino acid
      STRANDEDNESS:
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
US-09-338-876-9

  Query Match             2.9%;  Score 183;  DB 3;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-11;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        545 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 579
              |||||||||||||||||||||||||||||||||||
Db          1 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 35


RESULT 4
US-09-413-814-84
; Sequence 84, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 84
;   LENGTH: 757
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-84

  Query Match             2.1%;  Score 131;  DB 3;  Length 757;
  Best Local Similarity   21.1%;  Pred. No. 0.001;
  Matches  146;  Conservative   84;  Mismatches  251;  Indels  212;  Gaps   37;

Qy         54 VEKTFTSSLAVVSGART----TGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNL 109
              :|:|     ::|: ||      |  | |||                  ||:|||
Db        129 LERTLPRVASIVADARAEAALTSSAVAGIVAS------------------LPASAAA-- 167

Qy        110 TRACNAARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMC 169
                 || :|  :    ||   | :|  ||       | ||:
Db        168 -----AALQRLRWIATDGPS-PGPIEGPGA------ALRPESV---------------- 198

Qy        170 NVFLHY-----GGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVY 224
                 ||  |     |     | : ||::    |   :|     :   |||          | ::
Db        199 -AFLQYTSGSTGEPKGVMLTHGNL----LHNSRLIAHGFDLTSPDPV---------GVIW 244

Qy        225 PTPFYNTGLC--------HLIHDCVIAPMAVALR----VRNVTAV---ARGAAHLAFD-- 267
               |::: ||          |   :::|::   |     :| |:|:     |  : |:|
Db        245 LPPYHDMGLIGGILQALYRRIRVALMSPLSFLQRPMRWLRAVSALGASVSGGPNFAYDLC 304

Qy        268 ----ENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPSPSGGFERRLASIMA 323
                  | | |          :|      | | |  : ::  ::      || |    :
Db        305 VRKSSEEERAALD-------LRSWEVAFTGAEPVRADTLDRFARAFAVSGFRRE--AFYP 355

Qy        324 ADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRLNALGSYTARVAGVIGAMVF-- 381
                  | :| : |   | |           :   |    ||   |   |  || ||
Db        356 CYGLAEATLIVSGGARAEAPV----------LARLAPEEVELGRAVASAAE--GARVFVG 403

Qy        382 -----SPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLA----ANPQTDRDGHV 432
                  | :   :  |: :|     : |||      :  :|| :|      |:
Db        404 SGRALDPRA---VAIVDPAG----NELGPG-EIGEIW-VSGPSVAVGYWGRPE------- 447

Qy        433 LSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGT-- 490
                  |    : | : |  ||       ||   | |   ||      :| |
Db        448 ----------ETEATFGATLAGSAAPR------YLRTGDLGFLRGGE----LFVVGRSK 486

Qy        491 ----------FDSEIPCSLCEKH--TRPVC--AHTTVHRLRQRM-------PRFGQATRQ 529
                       | :|  ::   |   || |  | :  |  :|:        ||     |:
Db        487 DLIILRGRNHFPQDIEKTVESSHRAVRPGCSAAFSVEHEGEERLAVVCEVDPRVAADPRE 546

Qy        530 PIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQER 589
               :      : :::     |  :|  ||     :| : |   ::  || ||    |   ||
Db        547 IVAAREAVTAEHQ----LVAHAVALIAPGALPKTSSGKVRRRECRRAFLE----DALGER 598

Qy        590 -------LLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKI 642
                    ||| :|      :    : |: | :|| ||:    |  : :::       | :
Db        599 HVAFAPELLDDASPPDD---APPETEEPSGRSLLDALRS----TLARALRLDAGQIDDAL 651

Qy        643 ---REGL-SEATHSMALTFDPYSGAFCPITNFL 671
                  | || | |   :   |   :|   |:|: |
Db        652 PISRFGLDSLAAVELQHAFQVRTGRAIPLTSIL 684


RESULT 5
US-09-071-035-358
; Sequence 358, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  358:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1074 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-358
```

```
  Query Match           2.0%;  Score 123;  DB 4;  Length 1074;
  Best Local Similarity  18.3%;  Pred. No. 0.013;
  Matches  206;  Conservative  150;  Mismatches  415;  Indels  352;  Gaps   52;

Qy       250 RVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPS 309
Db        67 RTTSLYAEYNGAKQTVFC-IEPGVSIPTEVTHGY--------------------QKNPL 104

Qy       310 PSGGFERRLASIM--AADTALHAEVIFNTGIYEE-------------TPTDIKEWPMFI 353
Db       105 PSMSDKAKLVSVLWEKAGTDIDTNMVAQKMIWEEVNGYKLHSIKRLGGASVDIK------ 158

Qy       354 GMEGTLPRLNALGSYTAR-------VAGVIG-AMVFSPNSALYLTEVEDSGMTEAKDGGP 405
Db       159 SIEGKINK--AIEEYQKKPSFHNTTVKTILGQSTTLIDKNELNLSEFDKVVQNTA----- 211

Qy       406 GPSFNRFYQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLL 465
Db       212 ----NIDYRVIGNQLVLTP--------------NSNSKSGTLTLKKSAGTGTP-VAYKK 251

Qy       466 FYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQ 525
Db       252 AGLQTVMAGALDKPNTYAIKINVETKGS-LKIKKIDKESGDIVPETVFHL------DFGK 304

Qy       526 A------TRQPIGVFGTMNSQYSDCDPLGNYA---------PYLILRKPGDQTEAAKATM 570
Db       305 ALPSKDVTTDKDGI------SILDGIPHGTKVTITEKSVPDPYMIDTTPMAATIKAGETI 358

Qy       571 QDTYRATLERLFIDLEQERLLDRGAPCSSEGLS------SVIVDHPT---FRRILDTLRA 621
Db       359 SMTSKNMRQKGQILLEKTG-VETGTDLWNDNYSLAGNTFAIRKDSPAGEIVQEITTDEKG 417

Qy       622 RIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDP------YSGAFCPITNFLVKRT 675
Db       418 RAE--TPKELANALELGTYYVTE--TKSSNGFVNTFKPTKVELKYANQTVALVTSNVKGQ 473

Qy       676 HLAVVQDLAL----------SQCHCVFYGQQV--------EGRNFRNQFQPVLRRRFVDL 717
Db       474 NQEITGETTLTKEDKDTGNESQGKAEFKGAEYTLFTAKDGQAVKWSEAFK-------TEL 526

Qy       718 FNGGFISTRSITVTLSE-GPVSAPNPTLGQ------DAPAGRTFDGDLARVSVE------ 764
Db       527 VKGTKASDETVTLALDEKNQVAVKHLAINEYFWQETKAPEGYTLDETKYPVSIKKVDNNE 586

Qy       765 ----VIRDIRVKNRVV--------FSGNCTNLSEAA-------RARLVGLASAYQRQEKR 805
Db       587 KNAVITRDVTAKEQVIRFGFDFFKPAGSADGTAETGFNDLSFKVSPLEGTXEITGAEDKA 646

Qy       806 VDMLHGALGF--------------LLKQFHGLLFPRGM----PPNSKS---------PNP 838
Db       647 TTACNEQLGFDGYGKFENLPYGDYLLEEIEA---PEGFQKITPLEIRSTFKENKDDYAKS 703

Qy       839 QWFWTLLQRNQ--------MPADKLTHEEITTIAAVKRFTEEYAAINFINLP--PTCIGE 888
Db       704 EYVFTITEEGQKQPIKMVTVPYEKLTNNE---------FSVSLNRLMLYDLPEKEDSLTS 754

Qy       889 LAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKD---VTSAADIETQ 945
Db       755 LATWKDGNKKLNTLDFTE-LVDKL-------------RYNLHEIKEDWYVVAQAIDVEA- 799

Qy       946 AKALLEKTENLPELWTTAFTST-------------HLVRAAMNQRPMVVLGISISKYHGA 992
Db       800 TKAAQEKDEKAKPVVIAETTATLANKEKTGTWKILHKLTAEQ------VLDKSIVLFNYV 853

Qy       993 AGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSD 1052
Db       854 YENKVAFEAGNE--------------------------------PVA------KDASLNN 875

Qy      1053 QVRGIIVSGGAMVQLAIYATVVRAVGAR--------AQHMAFDDWLSLTDDEFLARDL-- 1102
Db       876 QAQ------------TVNCTIERHVSIQTKAHLEDGSQTFTHGDVMDMFDDVSVTHDVLD 923

Qy      1103 --EELHDQIIQTLETPWTVEGALEAVKILDE-------KTTAGD----GETPTNLAFNFD 1149
Db       924 GSKEAFETILYALLPDGTNKEIWKSGKIEHEVNDKEFTKTVLAEKVDTGKYPEGTKFTF- 982

Qy      1150 SCEPSHDTTSNV-------LNISGSNISGSTVPGLKRPPEDDE 1185
Db       983 -TEINYEKDGNVNGKHNEDLKEKSQTLTPKEVPTIPSTPKQPE 1024
```

RESULT 6
US-09-071-035-394
; Sequence 394, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 394:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1074 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-394

```
  Query Match           2.0%;  Score 123;  DB 4;  Length 1074;
  Best Local Similarity  18.3%;  Pred. No. 0.013;
  Matches  206;  Conservative  150;  Mismatches  415;  Indels  352;  Gaps   52;

Qy       250 RVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPS 309
Db        67 RTTSLYAEYNGAKQTVFC-IEPGVSIPTEVTHGY--------------------QKNPL 104

Qy       310 PSGGFERRLASIM--AADTALHAEVIFNTGIYEE-------------TPTDIKEWPMFI 353
Db       105 PSMSDKAKLVSVLWEKAGTDIDTNMVAQKMIWEEVNGYKLHSIKRLGGASVDIK------ 158

Qy       354 GMEGTLPRLNALGSYTAR-------VAGVIG-AMVFSPNSALYLTEVEDSGMTEAKDGGP 405
Db       159 SIEGKINK--AIEEYQKKPSFHNTTVKTILGQSTTLIDKNELNLSEFDKVVQNTA----- 211

Qy       406 GPSFNRFYQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLL 465
Db       212 ----NIDYRVIGNQLVLTP--------------NSNSKSGTLTLKKSAGTGTP-VAYKK 251

Qy       466 FYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQ 525
```

```
b       252 AGLQTVMAGALDKPNTYAIKINVETKGS-LKIKKIDKESGDIVPETVFHL------DFGK 304
y       526 A------TRQPIGVFGTMNSQYSDCDPLGNYA---------PYLILRKPGDQTEAAKATM 570
b       305 ALPSKDVTTDKDGI------SILDGIPHGTKVTITEKSVPDPYMIDTTPMAATIKAGETI 358
y       571 QDTYRATLERLFIDLEQERLLDRGAPCSSEGLS------SVIVDHPT---FRRILDTLRA 621
b       359 SMTSKNMRQKGQILLEKTG-VETGTDLWNDNYSLAGNTFAIRKDSPAGEIVQEITTDEKG 417
y       622 RIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDP------YSGAFCPITNFLVKRT 675
b       418 RAE--TPKELANALELGTYYVTE--TKSSNGFVNTFKPTKVELKYANQTVALVTSNVKGQ 473
y       676 HLAVVQDLAL----------SQCHCVFYGQQV--------EGRNFRNQFQPVLRRRFVDL 717
b       474 NQEITGETTLTKEDKDTGNESQGKAEFKGAEYTLFTAKDGQAVKWSEAFK-------TEL 526
y       718 FNGGFISTRSITVTLSE-GPVSAPNPTLGQ------DAPAGRTFDGDLARVSVE------ 764
b       527 VKGTKASDETVTLALDEKNQVAVKHLAINEYFWQETKAPEGYTLDETKYPVSIKKVDNNE 586
y       765 ----VIRDIRVKNRVV--------FSGNCTNLSEAA-------RARLVGLASAYQRQEKR 805
b       587 KNAVITRDVTAKEQVIRFGFDFFKFAGSADGTAETGFNDLSFKVSPLEGTXEITGAEDKA 646
y       806 VDMLHGALGF--------------LLKQFHGLLFPRGM----PPNSKS---------PNP 838
b       647 TTACNEQLGFDGYGKFENLPYGDYLLEEIEA---PEGFQKITPLEIRSTFKENKDDYAKS 703
y       839 QWFWTLLQRNQ--------MPADKLTHEEITTIAAVKRFTEEYAAINFINLP--PTCIGE 888
b       704 EYVFTITEEGQKQPIKMVTVPYEKLTNNE---------FSVSLNRLMLYDLPEKEDSLTS 754
y       889 LAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKD---VTSAADIETQ 945
b       755 LATWKDGNKKLNTLDFTE-LVDKL-------------RYNLHEIKEDWYVVAQAIDVEA- 799
y       946 AKALLEKTENLPELWTTAFTST-------------HLVRAAMNQRPMVVLGISISKYHGA 992
b       800 TKAAQEKDEKAKPVVIAETTATLANKEKTGTWKILHKLTAEQ------VLDKSIVLFNYV 853
y       993 AGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSD 1052
b       854 YENKVAFEAGNE--------------------------------PVA------KDASLNN 875
y      1053 QVRGIIVSGGAMVQLAIYATVVRAVGAR--------AQHMAFDDWLSLTDDEFLARDL-- 1102
b       876 QAQ------------TVNCTIERHVSIQTKAHLEDGSQTFTHGDVMDMFDDVSVTHDVLD 923
y      1103 --EELHDQIIQTLETPWTVEGALEAVKILDE-------KTTAGD----GETPTNLAFNFD 1149
b       924 GSKEAFETILYALLPDGTNKEIWKSGKIEHEVNDKEFTKTVLAEKVDTGKYPEGTKFTF- 982
y      1150 SCEPSHDTTSNV-------LNISGSNISGSTVPGLKRPPEDDE 1185
b       983 -TEINYEKDGNVNGKHNEDLKEKSQTLTPKEVPTIPSTPKQPE 1024
```

RESULT 7
US-09-252-991A-16913
; Sequence 16913, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16913
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16913

Query Match 1.9%; Score 118.5; DB 4; Length 370;
Best Local Similarity 20.2%; Pred. No. 0.0051;
Matches 93; Conservative 46; Mismatches 137; Indels 185; Gaps 19;

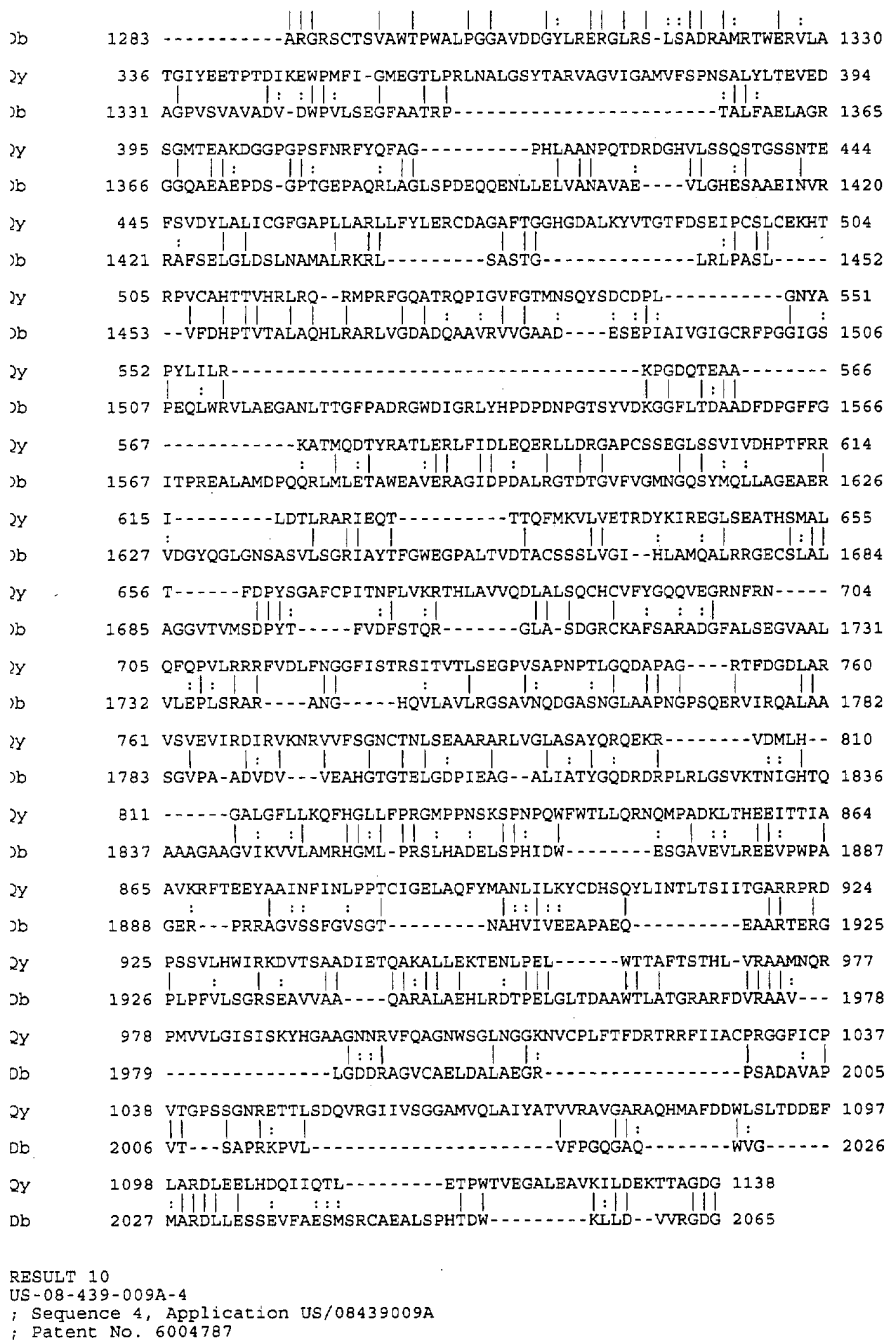
```
Qy      208 PDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAFD 267
Db       17 PDPCNAYH-----------------------------------------TDEQRWAAVLTRD 37

Qy      268 ENHEGA-VLPPDITYTYFQSSSSGTTTARGARRNDVN--STSKPSPSGGF--------ER 316
Db       38 AAADGAFVYAVKTTGVYCRPSSS----ARRPRRENVEFFATAEAAEAAGYRPSRRAAGDR 93

Qy      317 RLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGME------------GTLPRLNA 364
Db       94 RLAAEQRAERVAQACRMIETA---ETPPALEALAARLGMSPFHFHRLFKAETGLTPKAYA 150

Qy      365 LGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLAANP 424
Db      151 SAYRARRLRERLGQASASVTEAIY-----DSGFN---------SNSRFYESSSQRLGMRP 196

Qy      425 QTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGHGDAL 484
Db      197 RDYRDG-----------------------GAG----AAIRFAIGQCSLGAI-------- 220

Qy      485 KYVTGTFDSEIPCSLCEKHTRPVCA-------HTTVHRLRQRMPR---------FGQATR 528
Db      221 --------------LVAQSQRGICAILLGEEPEPLLRELQDQFPRAQLLGGDADFERLVA 266

Qy      529 QPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQ------DTYRATLERLF 582
Db      267 QVVG--------FVESPQLGLDLP-LDVRGTAFQERVWQALREIPPGSTASYAQIAERI- 316

Qy      583 IDLEQERLLDRGAP---------CSSEGLSSVIVDHPTFRR 614
Db      317 -----------GAPRAVRAVAQACAANRIAVAIPCHRVVRR 346
```

RESULT 8
US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-2

Query Match 1.8%; Score 111; DB 3; Length 5215;
Best Local Similarity 19.7%; Pred. No. 4.4;
Matches 250; Conservative 109; Mismatches 419; Indels 494; Gaps 66;

```
Qy       51 NLTVEKTFTSSLAVV--------SGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLP 102
Db      166 SLTVDAAQSSSLVAVHLACESLRAGESTTALV-AGVNL-----------------NILA 206
```

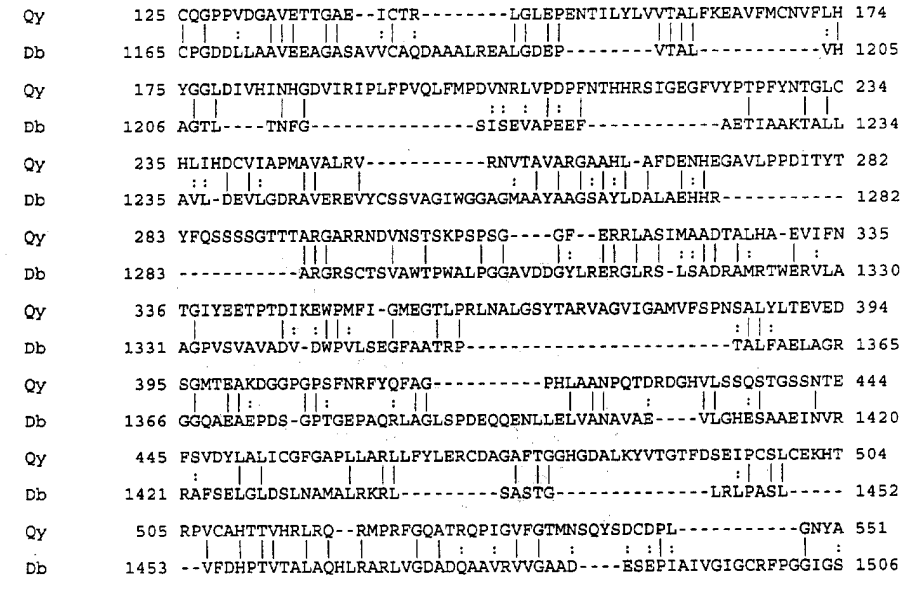
```
Qy      103 SSAAPNLTRACNAARERPGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALF 162
Db      207 ESA---------VTEERFG-----------------------GLSPDGTAYTFDARA-- 231

Qy      163 KEAVFMCNVFLHYGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGF 222
Db      232 -------NGFVRGEG-------GGVVVLKPL----------SRALADGDRVH------GV 261

Qy      223 VYPTPFYNTGLCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYT 282
Db      262 IRASAVNNDG----------------ATPGLTVPSRAAQEKVLREAYRKAALDPS-AVQ 303

Qy      283 YFQSSSSGTTTARGARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEET 342
Db      304 YVELHGTGT-----------PVGDPIEAAALGAVLGSARPADEPL---------LVGSA 342

Qy      343 PTDIKEWPMFIGMEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKD 402
Db      343 KTNVGHLEGAAGIVGLIKTLLALG--RRRIPASLNFRTPHPDIPL------DTLGLDVPD 394

Qy      403 GGPGPSFNRFYQFAGPHLAANPQTDRD----------GHVLSSQSTGSSNTEFSVD 448
Db      395 G---------------LREWPHPDRELLAGVSSFGMGGTNAHVVLSEGPAQGGEQPGID 438

Qy      449 YLALICGFGAPLLARLLFYLERCDAGA-----FTGGHGDALK------YVTGTFDSEI-P 496
Db      439 EETPV-----------------DSGAALPFVVTGRGGEALRAQARRLHEAVEADPELAP 480

Qy      497 CSLCEK--HTRPVCAHTTV------HRLRQRMPRFGQATRQPIGVFGTMNSQYSDCDPLG 548
Db      481 AALARSLVTTRTVFTHRSVVLAPDRARLLDGLGALAAGTPAPGVVTGT----------- 528

Qy      549 NYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSEGLSSVIVD 608
Db      529 --------PAPG----------------RLAVLF--------SGQGAQRTGMGM-ELYAA 555

Qy      609 HPTFRRILDTLRARIEQTTTQFMKVLVETRD----------YKIREGLSEATHSMALT 656
Db      556 HPAFATAFDAVAAELDPLLDRPLAELVAAGDTLDRTVHTQPALFAVEVALHRLVESWGVT 615

Qy      657 FDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVD 716
Db      616 PDLLAGH----SVGEISAAHVAGV--LSLRDA-----ARLVAARGRLMQALP-------- 656

Qy      717 LFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVV 776
Db      657 --EGGAM----VAVEASEEEVL---PHL-----AGRERELSLAAVN--------GPRAVV 694

Qy      777 FSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLF---------- 825
Db      695 LAG-----AERAVLDVAELLREQGRRTKRLSVSHA--------FHSPLMEPMLDDFRRVV 741

Qy      826 -------PR--------GMPPNSKSPNPQW----FWTLLQRNQMPADKLTHEEITTIAAV 866
Db      742 EELDFQEPRVDVVSTVTGLPVTA----GQWTDPEYW---------VDQV-RRPVRFLDAV 787

Qy      867 KRFTEEYAAINFINLPP--TCIGELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRD 924
Db      788 -RTLEESGADTFLELGPDGVCSAMAADSV----------RDQEAATAVSALRKG--RP-E 833

Qy      925 PSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGI 984
Db      834 PQSLLAALTTVFVRGHDVD---------------WTAAHGSTGTVRVPL---PTYAFQR 874

Qy      985 SISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSG 1044
Db      875 ERHWFDGAARTAAPLTAGR-SGTGAG--------------------------TGPAAG 905

Qy     1045 ---------------------NRETTLSDQVRGIIVSGGAMVQLAIYATVVRA-VGARA 1081
Db      906 VTSGEGEGEGEGAGAGGGDRPARHETT--ERVRAHVAA------VLEYDDPTRVELGLTF 957

Qy     1082 QHMAFDDWLS------LTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTA 1135
Db      958 KELGFDSLMSVELRNALVDDTGLRLPSGLLFDH-----PTP----RALAA--HLGDLLTG 1006

Qy     1136 GDGETPTNLAFNFDSCEPS--HDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDL---- 1189
Db     1007 GSGETG-----SADGIPPATPADTTAEPIAIIG---MACRYPGGVTSPED--LWRLVAEG 1056

Qy     1190 ----SGIPIKHG 1197
Db     1057 RDAVSGLPTDRG 1068


RESULT 9
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
;   APPLICANT:  Katz, L
;   APPLICANT:  Donadio, S
;   APPLICANT:  Mcalpine, J B
;   TITLE OF INVENTION:  Recombinant DNA Method for Producing
;   TITLE OF INVENTION:  Erythromycin Analogs
;   NUMBER OF SEQUENCES:  27
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Edward H. Gorman
;     STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;     STREET:  Park Rd
;     CITY:  Abbott Park
;     STATE:  IL
;     COUNTRY:  US
;     ZIP:  60064-3500
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/642,734C
;     FILING DATE:  17-JAN-91
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Danckers, Andreas M
;     REGISTRATION NUMBER:  32652
;     REFERENCE/DOCKET NUMBER:  4952.US.01
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  708-937-9396
;     TELEFAX:  708-938-2623
; INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  3567 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-07-642-734C-4

  Query Match             1.7%;  Score 106.5;  DB 2;  Length 3567;
  Best Local Similarity  19.9%;  Pred. No. 6.2;
  Matches  237;  Conservative  104;  Mismatches  384;  Indels  465;  Gaps   62;

Qy       125 CQGPPVDGAVETTGAE--ICTR--------LGLEPENTILYLVVTALFKEAVFMCNVFLH 174
Db      1165 CPGDDLLAAVEEAGASAVVCAQDAAALREALGDEP--------VTAL----------VH 1205

Qy       175 YGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLC 234
Db      1206 AGTL----TNFG--------------SISEVAPEEF----------AETIAAKTALL 1234

Qy       235 HLIHDCVIAPMAVALRV----------RNVTAVARGAAHL-AFDENHEGAVLPPDITYT 282
Db      1235 AVL-DEVLGDRAVEREVYCSSVAGIWGGAGMAAYAAGSAYLDALAEHHR---------- 1282

Qy       283 YFQSSSSGTTTARGARRNDVNSTSKPSPSG----GF--ERRLASIMAADTALHA-EVIFN 335
```



RESULT 10
US-08-439-009A-4
; Sequence 4, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:

; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-439-009A-4

Query Match 1.7%; Score 106.5; DB 3; Length 3567;
Best Local Similarity 19.9%; Pred. No. 6.2;
Matches 237; Conservative 104; Mismatches 384; Indels 465; Gaps 62;



```
Qy       552 PYLILR-----------------------------------KPGDQTEAA-------- 566
Db      1507 PEQLWRVLAEGANLTTGFPADRGWDIGRLYHPDPDNPGTSYVDKGGFLTDAADFDPGFFG 1566

Qy       567 ------------KATMQDTYRATLERLFIDLEQERLLDRGAPCSSEGLSSVIVDHPTFRR 614
Db      1567 ITPREALAMDPQQRLMLETAWEAVERAGIDPDALRGTDTGVFVGMNGQSYMQLLAGEAER 1626

Qy       615 I---------LDTLRARIEQT----------TTQFMKVLVETRDYKIREGLSEATHSMAL 655
Db      1627 VDGYQGLGNSASVLSGRIAYTFGWEGPALTVDTACSSSLVGI--HLAMQALRRGECSLAL 1684

Qy       656 T------FDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRN----- 704
Db      1685 AGGVTVMSDPYT-----FVDFSTQR-------GLA-SDGRCKAFSARADGFALSEGVAAL 1731

Qy       705 QFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAG----RTFDGDLAR 760
Db      1732 VLEPLSRAR----ANG-----HQVLAVLRGSAVNQDGASNGLAAPNGPSQERVIRQALAA 1782

Qy       761 VSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKR--------VDMLH-- 810
Db      1783 SGVPA-ADVDV---VEAHGTGTELGDPIEAG--ALIATYGQDRDRPLRLGSVKTNIGHTQ 1836

Qy       811 ------GALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIA 864
Db      1837 AAAGAAGVIKVVLAMRHGML-PRSLHADELSPHIDW--------ESGAVEVLREEVPWPA 1887

Qy       865 AVKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRD 924
Db      1888 GER---PRRAGVSSFGVSGT---------NAHVIVEEAPAEQ----------EAARTERG 1925

Qy       925 PSSVLHWIRKDVTSAADIETQAKALLEKTENLPEL------WTTAFTSTHL-VRAAMNQR 977
Db      1926 PLPFVLSGRSEAVVAA----QARALAEHLRDTPELGLTDAAWTLATGRARFDVRAAV--- 1978

Qy       978 PMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICP 1037
Db      1979 --------------LGDDRAGVCAELDALAEGR-----------------PSADAVAP 2005

Qy      1038 VTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGARAQHMAFDDWLSLTDDEF 1097
Db      2006 VT---SAPRKPVL----------------------VFPGQGAQ--------WVG------ 2026

Qy      1098 LARDLEELHDQIIQTL---------ETPWTVEGALEAVKILDEKTTAGDG 1138
Db      2027 MARDLLESSEVFAESMSRCAEALSPHTDW---------KLLD--VVRGDG 2065
```

RESULT 11
US-07-731-157A-7
; Sequence 7, Application US/07731157A
; Patent No. 5457032
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misset, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lenting, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/731,157A
; FILING DATE: 19910509
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90200962
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RAE-VENTER PH.D., BARBARA
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: GBRO-027/00US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas species
; STRAIN: SE83
US-07-731-157A-7

Query Match 1.7%; Score 106; DB 1; Length 774;
Best Local Similarity 20.2%; Pred. No. 0.4;
Matches 145; Conservative 79; Mismatches 251; Indels 242; Gaps 37;

```
Qy       258 ARGAAHLAFDENHEGAV-LPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPSPSGGFER 316
Db       183 AANALKLRYDDGGQDLLCIPPGVEAERLEADLAALRPAVDALLKAMGGDASDAAGGGSNN 242

Qy       317 ------RLAS---IMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRLNALGS 367
Db       243 WAVAPGRTATGRPILAGDPHRVFEI---PGMYAQHHLACDRFDM-IGL--TVP------- 289

Qy       368 YTARVAGVIGAMVFSPNSA--------------LYLTEVEDSGMTE-------------- 399
Db       290 ------GVPGFPHFAHNGKVAYCVTHAFMDIHDLYLEQFAEDGRTARFGNEPEPVAWRRD 343

Qy       400 --AKDGGPGPSFNRFYQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFG 457
Db       344 RIAVRGGADREFDIVETRHGPVIAGDP---LEGAALTLRSVQFAETDLSFDCLTRMP--G 398

Qy       458 APLLARLLFYLERCDAGAFTG--GH----GDALKYVTGTFDSEIPCSLCEKHTRPVCAHT 511
Db       399 ASTVAQLY------DATRGWGLIDHNLVAGDVAGSIGHLVRARVPSRPRENGWLPVPGWS 452

Qy       512 TVHRLR-----QRMPRFGQATRQPIGVFGTMNSQY----------SDCDPLGNYAPY--- 553
Db       453 GEHEWRGWIPHEAMPR---VIDPPGGLIVTANNRVVADDHPDYLCTDCHP-----PYRAE 504

Qy       554 ----LILRKPGDQTEAAKATMQDT-------YRATLERLFI--DLEQERL---------- 590
Db       505 RIMERLVASPAFAVDDAAAIHADTLSPHVGLLRARLEALGIQGSLPAEELRQTLIAWDGR 564

Qy       591 LDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEA- 649
Db       565 MDAGSQAASA--------YNAFRRAL--------------------TRLVTARSGLEQAI 596

Qy       650 THSMALT---FDPYSGAFCPITNFLVKRTHLAVVQ----DLALSQCHCVFYGQQVEGRNF 702
Db       597 AHPFAAVPPGVSPQGQVWWAVPT-LLRNDDAGMLKGWSWDEALSEALSV-ATQNLTGRGW 654

Qy       703 RNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAP------------NPTLGQDAP 749
Db       655 GEEHRP----RFTHPLSAQFPAWAALL-----NPVSRPIGGDGDTVLANGLVPSAGPEAT 705

Qy       750 AGRTFDGDLARVSVEVIRDIRVKNRVVFS-GNCTNLSEAARARLVGLASAYQRQEKRVDM 808
Db       706 YG------------------ALSRYVFDVGNWDN--------------------SRWVV 726
```

```
Qy        809 LHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW-----FWTLLQRNQMPADKLTHEEI 860
              ||| |              |:    |  |        |   ::: |: :| :|:
Db        727 FHGASG------------HPASPHYADQNAPWSDCAMVPMLYSWDRIAAEAVTSQEL 771


ESULT 12
S-08-541-780-7
 Sequence 7, Application US/08541780
 Patent No. 5935831
  GENERAL INFORMATION:
    APPLICANT:  Quax, Wilhelmus J.
    APPLICANT:  Misset, Onno
    APPLICANT:  Van der Laan, Jan M.
    APPLICANT:  Lenting, Herman B.M.
    TITLE OF INVENTION:  Mutated beta-lactam acylase genes
    NUMBER OF SEQUENCES:  50
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  COOLEY GODWARD CASTRO HUDDLESON & TATUM
      STREET:  FIVE PALO ALTO SQUARE, 4TH FLOOR
      CITY:  PALO ALTO
      STATE:  CALIFORNIA
      COUNTRY:  USA
      ZIP:  94306
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/541,780
      FILING DATE:
      CLASSIFICATION:  435
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  US/07/731,157
      FILING DATE:
      APPLICATION NUMBER:  EP 90200962
      FILING DATE:  18-APR-1990
    ATTORNEY/AGENT INFORMATION:
      NAME:  RAE-VENTER PH.D., BARBARA
      REGISTRATION NUMBER:  32,750
      REFERENCE/DOCKET NUMBER:  GBRO-027/00US
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  415-494-7622
      TELEFAX:  415-857-0663
      TELEX:  380816 COOLEY PA
  INFORMATION FOR SEQ ID NO:  7:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  774 amino acids
      TYPE:  amino acid
      TOPOLOGY:  linear
    MOLECULE TYPE:  protein
    HYPOTHETICAL:  NO
    ORIGINAL SOURCE:
      ORGANISM:  Pseudomonas species
      STRAIN:  SE83
JS-08-541-780-7

  Query Match             1.7%;  Score 106;  DB 2;  Length 774;
  Best Local Similarity   20.2%;  Pred. No. 0.4;
  Matches  145;  Conservative   79;  Mismatches  251;  Indels  242;  Gaps   37;

Qy        258 ARGAAHLAFDENHEGAV-LPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPSPSGGFER 316
              |  |  | :|:  :  : :|| :     ::  :    |  |    :   :  :  ||
Db        183 AANALKLRYDDGGQDLLCIPPGVEAERLEADLAALRPAVDALLKAMGGDASDAAGGGSNN 242

Qy        317 ------RLAS---IMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRLNALGS 367
                   | |:   |:| |     |:    |:| :       : | ||:  |:|
Db        243 WAVAPGRTATGRPILAGDPHRVFEI---PGMYAQHHLACDRFDM-IGL--TVP------- 289

Qy        368 YTARVAGVIGAMVFSPNSA-------------LYLTEVEDSGMTE-------------- 399
                 || |   |: |                 ||| :  : | |
Db        290 ------GVPGFPHFAHNGKVAYCVTHAFMDIHDLYLEQFAEDGRTARFGNEFEPVAWRRD 343

Qy        400 --AKDGGPGPSFNRFYQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFG 457
                |  ||   |:     || :| :|   :|  |: :|  : |:  | | |  :  |
Db        344 RIAVRGGADREFDIVETRHGPVIAGDP---LEGAALTLRSVQFAETDLSFDCLTRMP--G 398

Qy        458 APLLARLLFYLERCDAGAFTG--GH----GDALKYVTGTFDSEIPCSLCEKHTRPVCAHT 511
               |  :|:|       ||    |  |    ||    :     : :|   |   ||   :
Db        399 ASTVAQLY------DATRGWGLIDHNLVAGDVAGSIGHLVRARVPSRPRENGWLPVPGWS 452

Qy        512 TVHRLR-----QRMPRFGQATRQPIGVFGTMNSQY----------SDCDPLGNYAPY--- 553
                 |  |     : |||       | |:  | |::           :|| |     ||
Db        453 GEHEWRGWIPHEAMPR---VIDPPGGLIVTANNRVVADDHPDYLCTDCHP-----PYRAE 504

Qy        554 ----LILRKPGDQTEAAKATMQDT-------YRATLERLFI--DLEQERL---------- 590
                  ::  |    : | |  ||        || || | |   | | |
Db        505 RIMERLVASPAFAVDDAAAIHADTLSPHVGLLRARLEALGIQGSLPAEELRQTLIAWDGR 564

Qy        591 LDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEA- 649
              :| |:  :|         :  ||| |                ||    | || :|
Db        565 MDAGSQAASA--------YNAFRRAL-------------------TRLVTARSGLEQAI 596

Qy        650 THSMALT---FDPYSGAFCPITNPLVKRTHLAVVQ----DLALSQCHCVFYGQQVEGRNF 702
               |  |        |    :  :   |::     :::    | |||:    |   | : || :
Db        597 AHPFAAVPPGVSPQGQVWWAVPT-LLRNDDAGMLKGWSWDEALSEALSV-ATQNLTGRGW 654

Qy        703 RNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAP------------NPTLGQDAP 749
              : :|    ||    :  | :  ::       |||  |              |: | :|
Db        655 GEEHRP----RFTHPLSAQFPAWAALL-----NPVSRPIGGDGDTVLANGLVPSAGPEAT 705

Qy        750 AGRTFDGDLARVSVEVIRDIRVKNRVVFS-GNCTNLSEAARARLVGLASAYQRQEKRVDM 808
              |                          :| ||  ||  |                 |  :
Db        706 YG------------------ALSRYVFDVGNWDN--------------------SRWVV 726

Qy        809 LHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW-----FWTLLQRNQMPADKLTHEEI 860
              ||| |              |:    |  |        |   ::: |: :| :|:
Db        727 FHGASG------------HPASPHYADQNAPWSDCAMVPMLYSWDRIAAEAVTSQEL 771


RESULT 13
US-09-107-532A-3855
; Sequence 3855, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
```

; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3855:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1095 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1095
; SEQUENCE DESCRIPTION: SEQ ID NO: 3855:
US-09-107-532A-3855

Query Match 1.7%; Score 105.5; DB 4; Length 1095;
Best Local Similarity 18.9%; Pred. No. 0.86;
Matches 145; Conservative 112; Mismatches 298; Indels 211; Gaps 35;

```
Qy        551 APYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSEGLS------S 604
              |||  |    |   |  |   |:  | |::  :: |:   ::  |       :
Db        360 APYTIDTTPMTTTIKAGETIYVTSKNAQEKGQIILDKSG-VETGSDLWNDNYSLAGNTFA 418

Qy        605 VIVDHPTFRRILD-TLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDP---- 659
              :  | ||   : : |       | : :   :|   | : |  ::|:|    || |
Db        419 IRKDSPTGEIVQEMTTDENGHAETPKEIANALELGTYYVTE--TKASHGFVNTFKPVKVE 476

Qy        660 --YSGAFCPITNFLVKRTHLAVVQDLAL----------SQCHCVFYGQQV------EGR- 700
                |:    :    ||  :  |  :  |          :|   || | :       :|:
Db        477 LKYANQTVALVTSNVKGQNQEVTGETTLTKEDKDTGDKAQGKAVFEGTEYTLFTAKDGKA 536

Qy        701 -NFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAP-----NPTLGQD--APAGR 752
               :   |:|       ::  |  |  ::|: | |   :|      |    |:  || |
Db        537 VKWSEAFKP-------EMVKGTKASDETVTLALDEKNQAAVKHLAINEYYWQETKAPEGY 589

Qy        753 TFDGDLARVSVE----------VIRDIRVKNRVV--------FSGNCTNLSEAA------ 788
              | |     ||::         : ||:  | :::        |:|:    :|
Db        590 TLDETKYPVSIKKVDDNEKNAVITRDVTAKEQIIRFGFDFFKFAGSAAGTAETGFNDLTF 649

Qy        789 -RARLVGLASAYQRQEKRVDMLHGALGF-------------LLKQFHGLLFPRGM---- 829
               : | |       :::     :  |||              ||::      | |
Db        650 KVSPLEGTNEITGAEDEATTAYNEQLGFDGYGKFENLPYGDYLLEEVEA---PEGFQKIT 706

Qy        830 PPNSKSP---------NPQWFWTLLQRNQ--------MPADKLTHEEITTIAAVKRFTEE 872
              |   :|             :: :|: :::|         :| :|||:          | |:
Db        707 PLEIRSTFKENKEDFVKSEYVFTITEQDQKQPIKTVTVPYEKLTN---------KAFSVS 757

Qy        873 YAAINFINLP--PTCIGELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLH 930
               :   :||  :  ||:  |  |   | :: |:: |          |    ||
Db        758 LNRLMLYDLPEEEDSLTSLATWKDGNKELTSLDSTE-LVDKL------------SYNLH 803

Qy        931 WIRKD---VTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAMN-----QRPMV-- 980
              |::|   |  | |::   ||  || |    :   | ||  |          |  :
Db        804 EIKEDWYVVAQAIDVDA-TKAAQEKDEKAKPV-VIAETSATLANKEKTGTWKIQHKLTAE 861

Qy        981 -VLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVT 1039
               ||  :|  ::    |   |:||:                                ||
Db        862 QVLNKTIVLFNYVYENKEAFEAGD-------------------------------KPVA 889

Qy       1040 GPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGARAQHMAFDDWLSLTDDEFLA 1099
                 :: :|::| : :  :    | :   | :       :|      | : :|| :
Db        890 ------KDVSLNNQAQTVSCTVEHHVSIQTKAHLENG----SQTFTHGDVVDMFDDVSIT 939

Qy       1100 RDL----EELHDQIIQTLETPWTVEGALEAVKILDE-------KTT----AGDGETPTNL 1144
               |:    :|  : |: |   | :  :: ||  |        ||       |: |
Db        940 HDVLDGSKEAFETILYALLPDGTNKEIWKSGKIEYEVNDKEFTKTVLAKKVDTGKYPEGT 999

Qy       1145 AFNFDSCEPSHDTTSN-----VLNISGSNISGSTVPGLKRPPEDDE 1185
               | |       | | |       |      ::   || :   |: |
Db       1000 KFTFAEINYDKDGTINGKHNEDLKEKSQTLTPKEVPTILSTPKQPE 1045
```

RESULT 14
US-08-633-760-46
; Sequence 46, Application US/08633760
; Patent No. 5804429
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: FUJIMURA, TAKAO
; APPLICANT: ISHII, YOSHINORI
; APPLICANT: NOGUCHI, YUJI
; TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,760
; FILING DATE: 01-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-929-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-633-760-46

Query Match 1.7%; Score 105; DB 1; Length 774;
Best Local Similarity 20.7%; Pred. No. 0.5;
Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

```
Qy        109 LTR--ACNAARERFGFSRCQGPPVDGAVETTGAE-ICTR----LGLEPENTILYLVVTAL 161
              |||  |   | |  |    :    |  |   | | :| |     ||:| ::       :
Db         70 LTRRKALGRAAEWLG---AEAAEADILVRRLGMEKVCRRDFEALGVEAKD---------M 117

Qy        162 FKEAVFMCNVFLHYGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHH-----R 216
               :  |   | ||  |                || ||:  :        |:|:   |     |
Db        118 LRAYVAGVNAFLASGA-------------PL-PVEYGLLGAE---PEPWEPWHSIAVMR 159

Qy        217 SIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVALRVRNVTAV----ARGAAHLAFDENHEG 272
               :|                    :   ||  ::  : |:     |  |  | :|:
Db        160 RLG---------------------LLMGAVWFKLWRMLALPVVGAANALKLRYDDGGRD 197

Qy        273 AV-LPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPSPSGGFER------RLAS---IM 322
               : :||      ::  :    |  |    :  :  :  ||          | |:   |:
Db        198 LLCIPPGAEADRLEADLATLRPAVDALLKAMGGDASDAAGGGSNNWAVAPGRTATGRPIL 257

Qy        323 AADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRLNALGSYTARVAGVIGAMVFS 382
              | |      |:    |:| :       : | ||:  |:|            || | |:
Db        258 AGDPHRVFEI---PGMYAQHHLACDRFDM-IGL--TVP-------------GVPGFPHFA 298

Qy        383 PNSA-------------LYLTEVEDSGMTE---------------AKDGGPGPSFNRF 412
               |               ||| :    | |                |  ||    |:
Db        299 HNGKVAYCVTHAFMDIHDLYLEQFAGEGRTARFGNDFEPVAWSRDRIAVRGGADREFDIV 358
```

```
Qy        413 YQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCD 472
Db        359 ETRHGPVIAGDP---RDGAALTLRSVQFAETDLSFDCLTRMP--GASTVAQLY------D 407

Qy        473 AGAFTG--GH----GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLR-----QRMP 521
Db        408 ATRGWGLIDHNLVAGDVAGSIGHLVRARVPSRPRENGWLPVPGWSGEHEWRGWIPHEAMP 467

Qy        522 RFGQATRQPIGVFGTMNSQY----------SDCDPLGNYAPY-------LILRKPGDQTE 564
Db        468 R---VIDPPGGIIVTANNRVVADDHPDYLCTDCHP-----PYRAERIMKRLVANPAFAVD 519

Qy        565 AAKATMQDTYRATLERLFIDLEQERLLDRGA--PCSSEGLSSVIV--------------D 608
Db        520 DAAAIHADTLSP-----HVGLLRRRLEALGARDDSAAEGLRQMLVAWDGRMDAASEVASA 574

Qy        609 HPTFRRILDTL---RARIEQTTTQ-----------------FMKVLVETRDYKIREG--- 645
Db        575 YNAFRRALTRLVTDRSGLEQAISHPFAAVAPGVSPQGQVWWAVPTLLRDDDAGMLKGWSW 634

Qy        646 ---LSEA---------------------THSMALTFDPYSGAFCP 666
Db        635 DQALSEALSVASQNLTGRSWGEEHRPRFTHPLATQFPAWAGLLNP 679


RESULT 15
US-08-633-760-48
; Sequence 48, Application US/08633760
; Patent No. 5804429
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  SAITO, YOSHIMASA
;    APPLICANT:  FUJIMURA, TAKAO
;    APPLICANT:  ISHII, YOSHINORI
;    APPLICANT:  NOGUCHI, YUJI
;    TITLE OF INVENTION:  A NEW CEPHALOSPORIN C ACYLASE
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/633,760
;      FILING DATE:  01-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-929-0 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 774 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-633-760-48

 Query Match             1.7%;  Score 105;  DB 1;  Length 774;
 Best Local Similarity   20.7%;  Pred. No. 0.5;
 Matches  146;  Conservative   76;  Mismatches  241;  Indels  242;  Gaps   37;

Qy        109 LTR--ACNAARERFGFSRCQGPPVDGAVETTGAE-ICTR----LGLEPENTILYLVVTAL 161
Db         70 LTRRKALGRAAEWLG---AEAAEADILVRRLGMEKVCRRDFEALGVEAKD---------M 117

Qy        162 FKEAVFMCNVFLHYGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHH-----R 216
Db        118 LRAYVAGVNAFLASGA-------------PL-PVEYGLLGAE---PEPWEPWHSIAVMR 159

Qy        217 SIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVALRVRNVTAV----ARGAAHLAFDENHEG 272
Db        160 RLG---------------------LLMGSVWFKLWRMLALPVVGAANALKLRYDDGGRD 197

Qy        273 AV-LPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPSPSGGFER------RLAS---IM 322
Db        198 LLCIPPGAEADRLEADLATLRPAVDALLKAMGGDASDAAGGGSNNWAVAPGRTATGRPIL 257

Qy        323 AADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRLNALGSYTARVAGVIGAMVFS 382
Db        258 AGDPHRVFEI---PGIYAQHHLACDRFDM-IGL--TVP------------GVPGFPHFA 298

Qy        383 PNSA-------------LYLTEVEDSGMTE---------------AKDGGPGPSFNRF 412
Db        299 HNGKVAYCVTHAFMDIHDLYLEQFAGEGRTARFGNDFEPVAWSRDRIAVRGGADREFDIV 358

Qy        413 YQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCD 472
Db        359 KTRHGPVIAGDP---RDGAALTLRSVQFAETDLSFDCLTRMP--GASTVAQLY------D 407

Qy        473 AGAFTG--GH----GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLR-----QRMP 521
Db        408 ATRGWGLIDHNLVAGDVAGSIGHLVRARVPSRPRENGWLPVPGWSGEHEWRGWIPHEAMP 467

Qy        522 RFGQATRQPIGVFGTMNSQY----------SDCDPLGNYAPY-------LILRKPGDQTE 564
Db        468 R---VIDPPGGIIVTANNRVVADDHPDYLCTDCHP-----PYRAERIMKRLVANPAFAVD 519

Qy        565 AAKATMQDTYRATLERLFIDLEQERLLDRGA--PCSSEGLSSVIV--------------D 608
Db        520 DAAAIHADTLSP-----HVGLLRRRLEALGARDDSAAEGLRQMLVAWDGRMDAASEVASA 574

Qy        609 HPTFRRILDTL---RARIEQTTTQ-----------------FMKVLVETRDYKIREG--- 645
Db        575 YNAFRRALTRLVTDRSGLEQAISHPFAAVAPGVSPQGQVWWAVPTLLRDDDAGMLKGWSW 634

Qy        646 ---LSEA---------------------THSMALTFDPYSGAFCP 666
Db        635 DQALSEALSVASQNLTGRSWGEEHRPRFTHPLATQFPAWAGLLNP 679


Search completed: January 30, 2004, 13:18:32
Job time : 27 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:10:12 ; Search time 46 Seconds
(without alignments)
6748.637 Million cell updates/sec

Title: US-09-769-699-2
Perfect score: 6294
Sequence: 1 MENTQKTVTVPTGPLGYVYA.....DELFDLSGIPIKHGNITMEM 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4627	73.5	1194	12	Q9E1Y7	Q9e1y7 cercopithec
2	3583	56.9	1208	12	O39273	O39273 equine herp
3	3371.5	53.6	1203	12	Q89549	Q89549 bovine herp
4	3327.5	52.9	1177	12	O92611	O92611 pseudorabie
5	3012	47.9	1197	12	Q69101	Q69101 herpes simp
6	2796.5	44.4	1190	12	Q9PZ52	Q9pz52 turkey herp
7	2793.5	44.4	1190	12	Q9QTB9	Q9qtb9 marek's dis
8	2788	44.3	1191	12	Q9E6P0	Q9e6p0 turkey herp
9	2781.5	44.2	1190	12	Q9DPQ8	Q9dpq8 meleagrid h
10	2764.5	43.9	1190	12	Q9E1G3	Q9e1g3 meleagrid h
11	2710	43.1	1191	12	Q9IBU6	Q9ibu6 turkey herp
12	1445	23.0	999	12	Q9QH63	Q9qh63 gallid herp
13	683.5	10.9	1132	12	O40913	O40913 kaposi's sa
14	675	10.7	1133	12	P88904	P88904 kaposi's sa
15	652.5	10.4	1103	12	Q992Z6	Q992z6 murid herpe
16	651.5	10.4	1103	12	O41928	O41928 murid herpe
17	651.5	10.4	1127	12	Q993K9	Q993k9 callitrichi
18	645	10.2	1125	12	Q8B414	Q8b414 porcine lym
19	644	10.2	1125	12	Q8B408	Q8b408 porcine lym
20	633	10.1	1128	12	Q9YTQ7	Q9ytq7 ateline her
21	623.5	9.9	1133	12	Q8UZD2	Q8uzd2 cercopithic
22	620	9.9	1132	12	Q9WRU1	Q9wru1 macaca mula
23	618	9.8	1145	12	Q66611	Q66611 equine herp
24	596	9.5	1127	12	O36360	O36360 alcelaphine
25	589.5	9.4	1126	12	Q8JYD5	Q8jyd5 porcine lym
26	576.5	9.2	1142	12	Q8B424	Q8b424 porcine lym
27	576.5	9.2	1142	12	Q8B3X5	Q8b3x5 porcine lym
28	561.5	8.9	1134	12	Q99D22	Q99d22 bovine herp
29	505	8.0	1223	12	Q8QS31	Q8qs31 chimpanzee
30	484	7.7	1131	12	O56282	O56282 human herpe
31	401	6.4	1281	12	Q85425	Q85425 rat cytomeg
32	237.5	3.8	483	12	Q82172	Q82172 leporid her
33	235	3.7	84	12	O56864	O56864 feline herp
34	156.5	2.5	1841	2	Q9FB25	Q9fb25 streptomyce
35	143.5	2.3	178	12	Q64908	Q64908 alcelaphine
36	138	2.2	1869	2	Q8L334	Q8l334 aphanizomen
37	137	2.2	1504	2	Q9ZGA6	Q9zga6 streptomyce
38	129.5	2.1	919	5	Q9VA02	Q9va02 drosophila
39	128.5	2.0	919	5	Q8IGV6	Q8igv6 drosophila
40	123.5	2.0	1287	16	Q8EV67	Q8ev67 mycoplasma
41	120.5	1.9	788	16	Q8U9K3	Q8u9k3 agrobacteri
42	120.5	1.9	968	4	Q13025	Q13025 homo sapien
43	120.5	1.9	1555	5	Q8T1Q4	Q8t1q4 trypanosoma
44	120	1.9	2152	16	Q9EX54	Q9ex54 streptomyce
45	119.5	1.9	701	16	Q92EA7	Q92ea7 listeria in

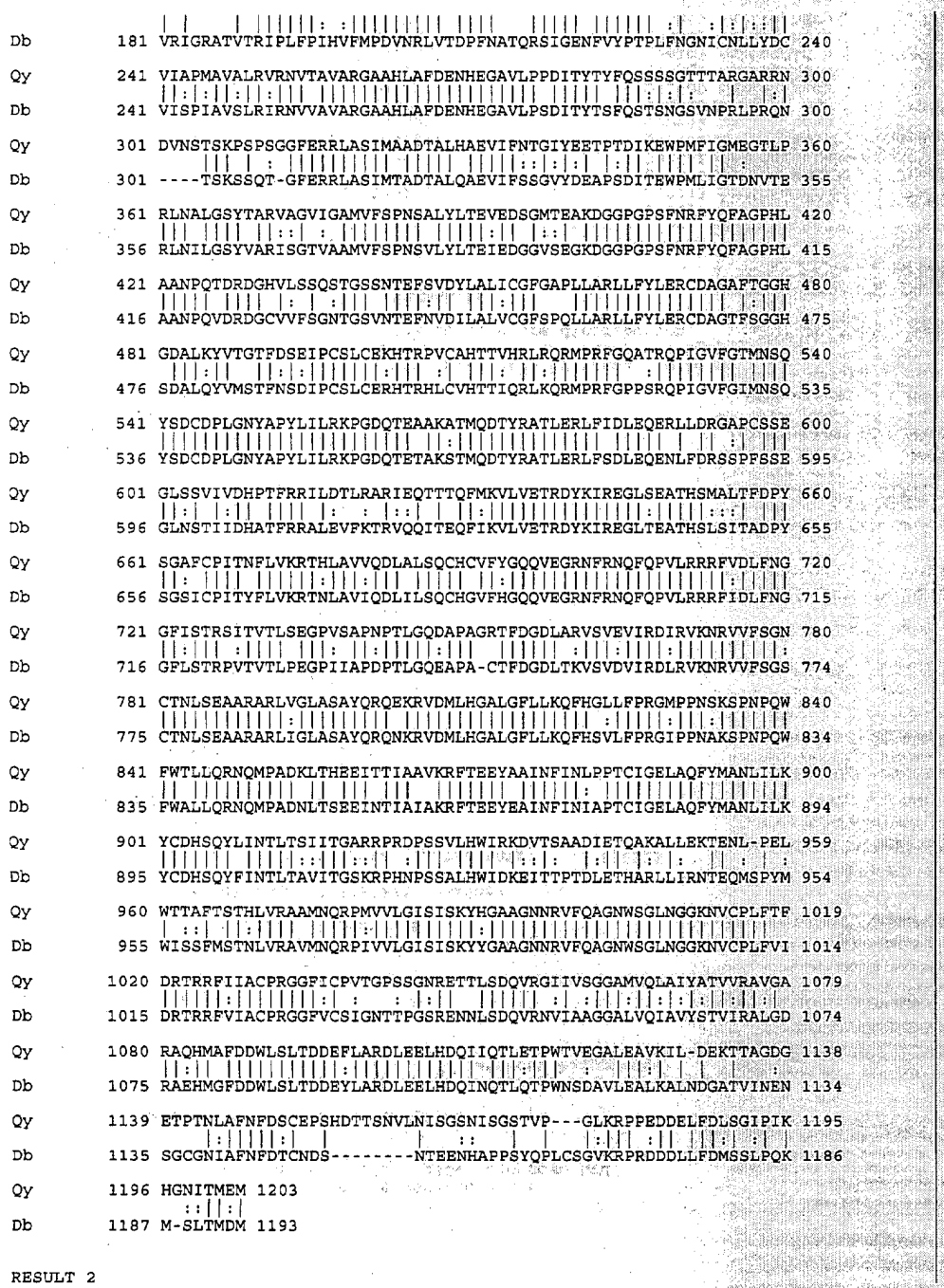
ALIGNMENTS

RESULT 1
Q9E1Y7
ID Q9E1Y7 PRELIMINARY; PRT; 1194 AA.
AC Q9E1Y7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SsDNA binding protein.
OS Cercopithecine herpesvirus 7.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=35245;
RN [1]
RP SEQUENCE FROM N.A.
RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
RT "Complete Sequence of the Simian Varicella Virus Genome.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF275348; AAG27202.1; -.
DR InterPro; IPR000635; Viral_DNA_bind.
DR Pfam; PF00747; viral_DNA_bp; 1.
SQ SEQUENCE 1194 AA; 131968 MW; EBA7F3C841965897 CRC64;

Query Match 73.5%; Score 4627; DB 12; Length 1194;
Best Local Similarity 71.8%; Pred. No. 0;
Matches 867; Conservative 143; Mismatches 178; Indels 20; Gaps 8;

```
Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
              || ||||:|||||||||||||  ||||| ||| |:||| |||||||:||:|||||||:|||
Db          1 MEATQKTITVPTGPLGYVYACNKEDLDLPEISLLSARSIDSDLALIPLIRNLTVEKSFTS 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
              ||||||| :|||||||:|||| |||:||: |||||| |: ||| :||||:||| |||||
Db         61 SLAVVSGTKTTGLAGAGVTLKLITSHYYPTAFVFHGGHHISPSSFSPNLTQACNTARERF 120

Qy        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
              ||   :| |||||:|||||:|| |: | ||:  |||:||:||||||:||| ||:|||:|
Db        121 GFMNKRGSPVDGAIETTGADICNRIHLNPEHAFLYLIVTSLFKEAVYMCNSFLYYGGIDN 180

Qy        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
```

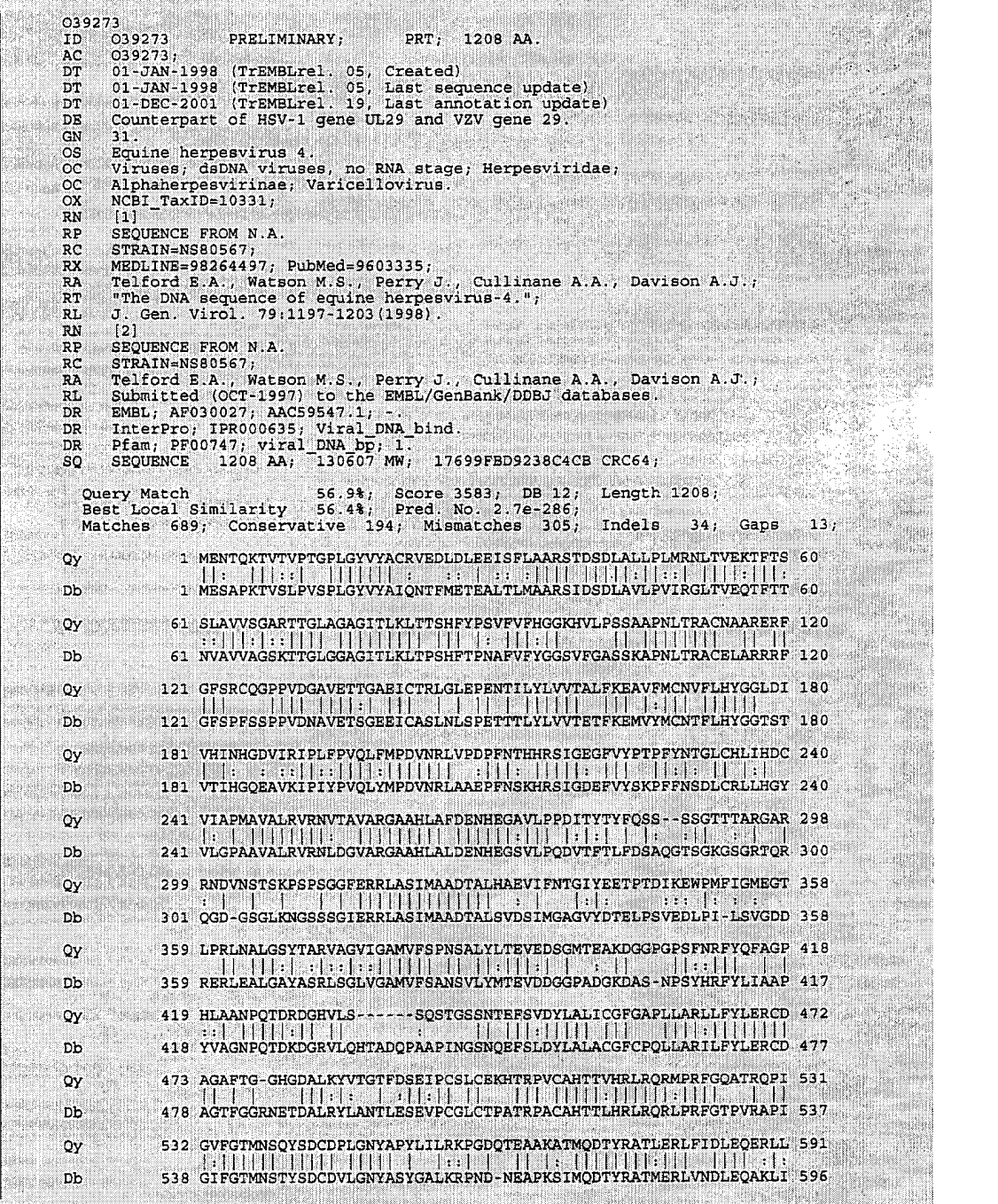
09/769,699 2-2-04
Search Notes

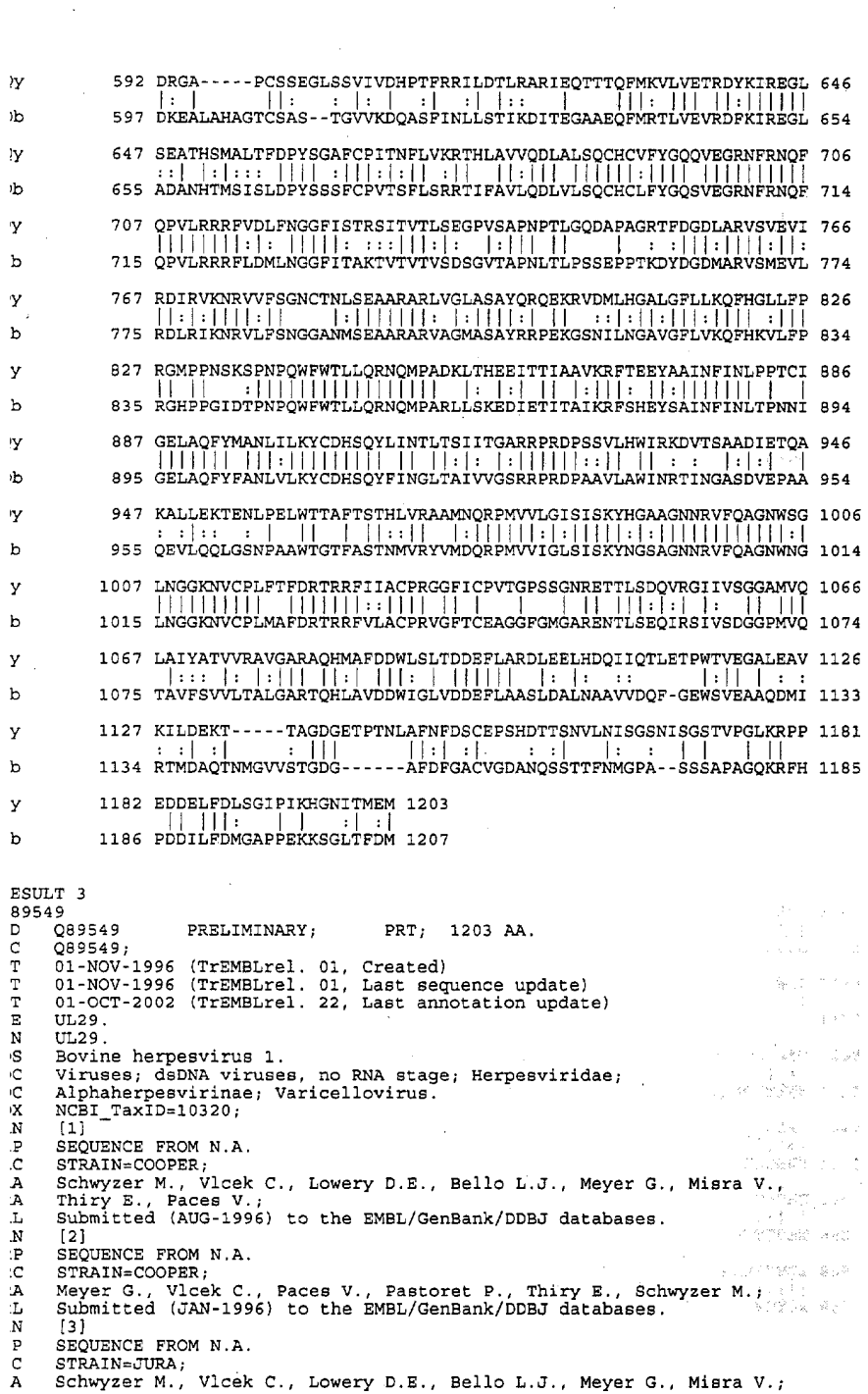


RESULT 2

O39273
ID O39273 PRELIMINARY; PRT; 1208 AA.
AC O39273;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Counterpart of HSV-1 gene UL29 and VZV gene 29.
GN 31.
OS Equine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10331;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RX MEDLINE=98264497; PubMed=9603335;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-4.";
RL J. Gen. Virol. 79:1197-1203(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF030027; AAC59547.1; -.
DR InterPro; IPR000635; Viral_DNA_bind.
DR Pfam; PF00747; viral_DNA_bp; 1.
SQ SEQUENCE 1208 AA; 130607 MW; 17699FBD9238C4CB CRC64;

Query Match 56.9%; Score 3583; DB 12; Length 1208;
Best Local Similarity 56.4%; Pred. No. 2.7e-286;
Matches 689; Conservative 194; Mismatches 305; Indels 34; Gaps 13;





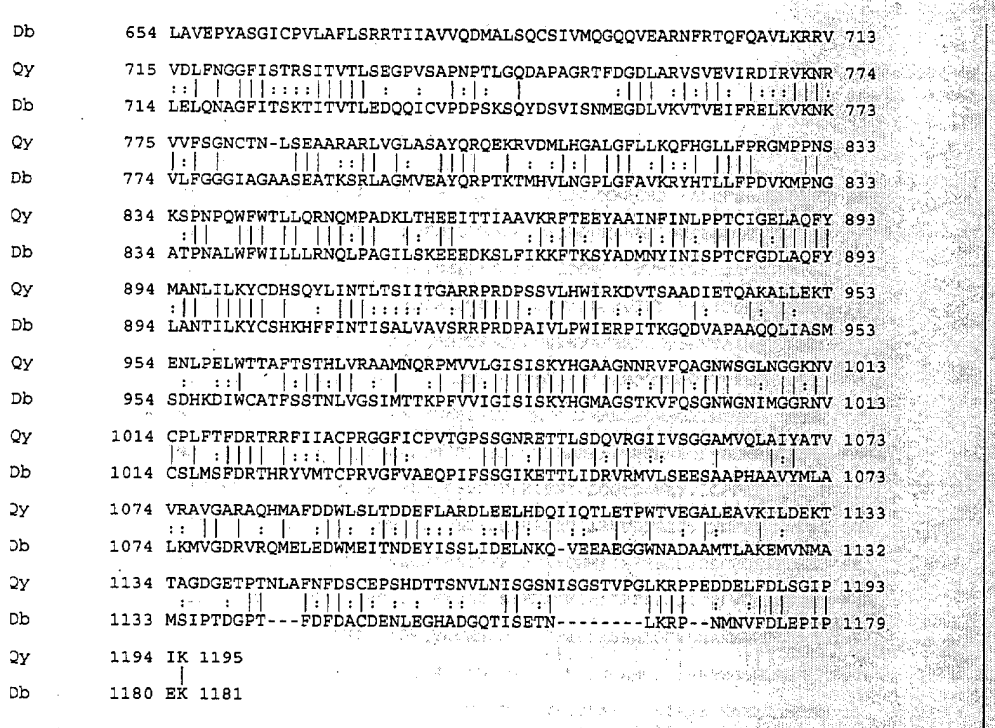
ESULT 3
89549
D Q89549 PRELIMINARY; PRT; 1203 AA.
C Q89549;
T 01-NOV-1996 (TrEMBLrel. 01, Created)
T 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
T 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
E UL29.
N UL29.
S Bovine herpesvirus 1.
C Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Varicellovirus.
X NCBI_TaxID=10320;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=COOPER;
A Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
A Thiry E., Paces V.;
L Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
N [2]
P SEQUENCE FROM N.A.
C STRAIN=COOPER;
A Meyer G., Vlcek C., Paces V., Pastoret P., Thiry E., Schwyzer M.;
L Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
N [3]
P SEQUENCE FROM N.A.
C STRAIN=JURA;
A Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.;
L Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
N [4]
P SEQUENCE FROM N.A.
C STRAIN=JURA;
A Schwyzer M.;
L Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
R EMBL; Z78205; CAB01596.1; -.
R EMBL; X94677; CAA64336.1; -.
R EMBL; AJ004801; CAA06104.1; -.
R InterPro; IPR000635; Viral_DNA_bind.
R Pfam; PF00747; viral_DNA_bp; 1.
Q SEQUENCE 1203 AA; 127409 MW; 8299D64966A9654F CRC64;

```
Qy        833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
              | :|||||||||||||||||   |: |:| || |||||:: |||||:||  |  : |||||
Db        837 SPTPNPQWFWTLLQRNQMPARLLSKEDIETIGAVKRFSDAYAAINYVNLTPGNVAELAQF 896

Qy        893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
              |||||:|::||| |: || ||:|: ||||||||::|:||:|: :  | |     |: :| :
Db        897 YMANLVLRHCDHKQFYINGLTAIVAGARRPRDPAAVMHWVRRRIADAPDAAAAAEEVLRE 956

Qy        953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
              |   ||:|    | :::|||: |   || ||||:|||||:|:|||||||||||||||||||
Db        957 AEARPEVWAGTFAASNLVRSVMATRPAVVLGLSISKYNGSAGNNRVFQAGNWSGLNGGKN 1016

Qy       1013 VCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGII-VSGGAMVQLAIYA 1071
              ||||  ||:||||::|||| ||:|    |    ||: || : :| ||  | ||: | |:|
Db       1017 VCPLLCFDKTRRFVLACPRAGFVCATAGAGGANRDNTLVEALREIINGSNGALAQTAVYG 1076

Qy       1072 TVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETP--WTVEGALEAVKIL 1129
              |::|:| | :||  ||| :| :||| |: : || :   :    |  |: ||| : :: |
Db       1077 AVLQALGPRVEHMDLDDWAALVEDEFFAQSMVELTE---RAAARPGGWSPEGAADMLREL 1133

Qy       1130 DEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLN--ISGSNISGSTVPGLKRPPEDDELF 1187
              | | : | |   ||:| :|  :    :      ::|   :|   || || |: ||||
Db       1134 -ELEAAAEAE-PAGGAFDFGACASAAPDATYAFGGPVAG---AGPAAPGAKR-PDLDELF 1187

Qy       1188 DLSGIPIKHGNITMEM 1203
              :: |:| |    :|::|
Db       1188 EM-GVPDKRPALTIDM 1202


RESULT 4
O92611
ID   O92611      PRELIMINARY;      PRT;  1177 AA.
AC   O92611;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   DNA-binding protein.
GN   DBP.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10345;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TNL;
RX   MEDLINE=98455382; PubMed=9784061;
RA   Wu S.-L., Hsiang C.-Y., Ho T.-Y., Chang T.-J.;
RT   "Identification, expression, and characterization of the pseudorabies
RT   virus DNA-binding protein gene and gene product.";
RL   Virus Res. 56:1-9(1998).
DR   EMBL; U80909; AAC63429.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding.
SQ   SEQUENCE   1177 AA;  125408 MW;  BA87AF9CFC961707 CRC64;

  Query Match             52.9%;  Score 3327.5;  DB 12;  Length 1177;
  Best Local Similarity   53.5%;  Pred. No. 3.4e-265;
  Matches  655;  Conservative  182;  Mismatches  318;  Indels   69;  Gaps   17;

Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
              ||    |||||   ||||||   :| |  : :: | ||| |    |: ||:| |||  |
Db          1 MEAAAKTVTVRAAPLGYVYVTPIEALRRDLLALLVARSADDAAAVAPLVRGLTVEAGFAG 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
                :|||:|||||||  | |:|||| :||:|:|| || |  | |||||| |:||| ||| ||
Db         61 HVAVVAGARTTGL-GGGLTLKLTPNHFHPNVFFFHNGDCVPPSSAAPALSRACEAARARF 119

Qy        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
              |||   :  ||| | |||||:| ||||  :    ||||   |||||::|| ||||||
Db        120 GFSAYR-TPVDNAEETTGAEVCARLGLAADAHAAYLVVADGPKEAVYLCNAFLHYGGAGT 178

Qy        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
              |  ||  |||  |  |  :|  :|  |  |  |  |  |  |  |  |  |  |  |
Db        179 VSINGHEAWRVPLYPVHLFMPDVNRLVADPFNAKNRSISEEFVYPRPFFNGPLCRLLHGY 238

Qy        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSSGTTTARGARRN 300
              |:| ||| |||| |||||||||||||||| |||| |:|  |   |  |  |
Db        239 VLGPAAVATRVRNLDAVARGAAHLAFDENHESAVLPADVTFTLFEQRRRGGDAA------ 292

Qy        301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMF-------- 352
                  :|| |||:||:|:|| ||  |  :   |:|:| |  : :|||
Db        293 ----------AGGLERRMASVMSADAALSLEALVAAGVYDEEPPALDDWPVLSEAGAKDG 342

Qy        353 ----IGMEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPS 408
                  :  |||:|  :| ||:|||| |||| |||| |||||:|:|  |  :|  | |||
Db        343 GAAAASAGAPVSQAAALGAYVSRAAGLVGALVFSSNSVLYLTEVDDAGAADARKEGAGPS 402

Qy        409 FNRFYQFAGPHLAANPQTDRDGHVL----SSQSTGSSNTEFSVDYLALICGFGAPLLARL 464
              |||||| | :|| |||||:|| ||    |  :|  | :|:::|: |||   ||||:
Db        403 FNRFYQVAAPYLAGNPQTDKDGRVLLHTASQPATAPGNHDFAMDHLVMACGFCPQLLARV 462

Qy        465 LFYLERCDAGAFTG-GHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRF 523
              |||||||| | |  |::|  |: |:|||||::  :|  ||||:|||| |
Db        463 LFYLERCDAGTFVGRGDVDAVRYVAGSLDAEVPCSLCDRASRPACAHTTLHRLRHR-SRA 521

Qy        524 GQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFI 583
                 | :|||||||| |||| |||| |  |::| | :|  || |||||| :|:||
Db        522 SARRRSPMGVFGTMNSAYSDCDVLGNYASYGALKRPND-SEPPKAIMQDTYRAAVDRLLA 580

Qy        584 DLEQERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIR 643
              |:   |          :  :|| || |  || :||   :|:: |||||:|:|
Db        581 DVAGAR-----------IGETVTDHAGFRHALRALRDTVEQAADRFVRTLVETRDFKLR 628

Qy        644 EGLSEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFR 703
              : | :| :|:|: |||||| || :|:| :|| |||:|||||||| | :|| |||||||
Db        629 DALYDANHTMSLSLDPYSGALCPATSFLARRTLLAVLQDLALSQCHGVLHGQPVEGRNFR 688

Qy        704 NQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSV 763
              |||||||||| || ||||:: :::|||||:| : ||:|| |    |  | ||||||||
Db        689 NQFQPVLRRRVVDMLNGGFVTAKTVTVTLADG-IVAPDPTKGSAEPPARDHDGDLARVSF 747

Qy        764 EVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGL 823
              ||:|::|||:||:||    :||:|||||:|||| |||:|  |||:|  |||||| |
Db        748 EVLRELRVKSRVMFSTGSGSLSDAARARVAGLAGAYQRPDTAVDVLNGPLGFLLKQHHAT 807

Qy        824 LFPRGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPP 883
              ||||| ||  :|||||||||||||||| || :: |||||||||: :| ||||:|| |
Db        808 LFPRGKPPGGQSPNPQWFWTLLQRNQLPARLLTKDDIETIAAVKRFSVDYGAINYVNLTP 867

Qy        884 TCIGELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIE 943
              : ||||||:|||||:|||| |: ||:||   :|||||:|:|:|:: :  | |
Db        868 GTVAELAQFYLANLILRYCDHKQFFINSLTGITMQSKRPRDPAAVMAWVRRPLADAADAE 927

Qy        944 TQAKALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGN 1003
              |: :|:    :  : |  :||:|:|: |  ||:|||||: :|||| ||||||||||
Db        928 RAARFVLDAPRD--DTWVATYTSSHLLRSVMASRPLVVLGLGVSKYHGMAGNNRVFQAGN 985

Qy       1004 WSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGA 1063
              |||||||:|||| |||||  |::|||| || |  ||  :|  :|  :|  :|  ||
Db        986 WSGLNGGKHVCPLMVFDRTRHFVLACPRVGFTCSQTGGGAGLHDHSLGEHVKTILADGGP 1045

Qy       1064 MVQLAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL 1123
              :|| |:|| |: |:||| ||:  ||| :: ||||||  | |:: ::    :  |:|| |
Db       1046 LVQTAVYAAVLHALGARTQHLEPDDWRAIVDDEFLAAALAEINGRVADR-DGRWSVEAAA 1104

Qy       1124 EAVKILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNIS-GSTVP---GLKR 1179
              | |: : :  |  |  |  | ||:| :|  |  |  |  :| :: |  |  | ||
Db       1105 ELVRDLEGQTGADGGGEET--AFDFGACGAGGD--------AGAGLAPASLAPAELGGKR 1154

Qy       1180 PPEDDELFDLSGIPIKHGNITMEM 1203
              || :|:|||:  |  |  :| :|
Db       1155 PPPEDDLFDMGAPPEK--RLTFDM 1176


RESULT 5
```

```
)69101
D   Q69101      PRELIMINARY;      PRT;  1197 AA.
AC   Q69101;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DNA binding protein ICP8.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KN;
RX   MEDLINE=93228441; PubMed=8385914;
RA   Toh Y., Tanaka S., Liu Y., Mori R.;
RT   "Nucleotide sequence of the major DNA binding protein of herpes
RT   simplex virus type 2 and comparison with the type 1 counterpart.";
RL   Arch. Virol. 129:183-196(1993).
DR   EMBL; D10658; BAA01507.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
SQ   SEQUENCE   1197 AA;  128470 MW;  AA3ADA75B8865BFE CRC64;

 Query Match            47.9%;  Score 3012;  DB 12;  Length 1197;
 Best Local Similarity  49.0%;  Pred. No. 4e-239;
 Matches  595;  Conservative  212;  Mismatches  368;  Indels   40;  Gaps   16;

Qy          4 TQKTVTVPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
              |  || || ||:|||   ||  | |:|  :| |:||| |:|:|: ||:  ||||  | ::
Db          7 TTTTVKVPPGPMGYVYGRACPAEGLEL--LSLLSARSGDADVAVAPLIVGLTVESGFEAN 64

Qy         62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFG 121
              :| | |: |||| |  ::|||  ||: |||:|||||:|: ||: |||||| |  ||  ||
Db         65 VAAVVGSGTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQAPNLTRLCERARRHFG 124

Qy        122 FSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIV 181
              ||    | |   |||  :| ||||:|:  :||||:|  |:||| : | ||| ||:| |
Db        125 FSDYAPRPCDLKHETTGDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGGMDKV 184

Qy        182 HINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCV 241
              |   :| |||::|:|:|||| :|:: |||| :|||||| | || ||:|  |  |: : |
Db        185 TIGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAV 244

Qy        242 IAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRND 301
              : | ||||| ||| |||| ||||||||||||| || |||:| |::|       ||||
Db        245 VGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KPQRGAR--- 299

Qy        302 VNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPR 361
                  |:||||:||||:|| | ||  | | :  :::| | ||  ||:  | |    |
Db        300 --DAGNKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPPDITTWPLLEGQETPAAR 357

Qy        362 LNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLA 421
              |:|:| || ||::|||||| ||||:||||:|:|  : ||    ||| ||:   | |:|
Db        358 AGAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVA 416

Qy        422 ANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAG 474
              |||| ||:|||:       ::   | :  ||: ::||::|||   |||::|||||||| |
Db        417 ANPQLDREGHVVPGYEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYLERCDGG 475

Qy        475 AFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGV 533
              |     |  :|  :  :::||:||    ||  ||||| ||| | |:|  | |  |||
Db        476 VIVGRQEMDVFRYVADSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGV 535

Qy        534 FGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDR 593
              |||||| ||||| ||||| :  |::  | :|  :  ||:|||| ||:  :|| : :|:
Db        536 FGTMNSAYSDCDVLGNYAAFSALKR-ADGSENTRTIMQETYRAATERVMAELEALQYVDQ 594

Qy        594 GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSM 653
              | :   | ::| :      ::: ::  :::   | |: |:| |::| |:||:|| |:|
Db        595 AVPTALGRLETIIGNREALHTVVNNIKQLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654

Qy        654 ALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRR 713
              :|: |||:   ||:   | :|::||| |||||||| || || |||||||||||||||||
Db        655 SLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRRR 714

Qy        714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
              :|||| ||:| :::|| |||  : ||: | || |||  :|:||:|||::   :::|||
Db        715 VMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVK 774

Qy        773 NRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPN 832
              :||:|:|   | ||||:||:  | |||| :||||:| | |||||||| ::|| | ||
Db        775 SRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPPG 834

Qy        833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
              |  |||||||| |||||:||  |: |:| ||| :|||: :| |||||||  |  : ||| :
Db        835 SNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMY 894

Qy        893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
              |||| ||:|||| | |||||::| |:|||    :   | :     | :|  |:||::
Db        895 YMANQILRYCDHSTYFINTLTAVIAGSRRPPSVQAAAAWAPQ---GGAGLEAGARALMDS 951

Qy        953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
              :  |  ||: | | :|:|  | |||||||:|||||:| |||:|||||||: | ||||
Db        952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGKN 1011

Qy       1013 VCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYAT 1072
              |||  ||||:|::|||| ||:|  :    |  | :| :|:|||| |||  | :::
Db       1012 ACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071

Qy       1073 VVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVKI 1128
              |:::| | | :  :|||:| :||:|: :: |   : ::     |: : ||    ||  :
Db       1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEAL 1131

Qy       1129 LDEKTTAGDGETPTNLAFNPDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFD 1188
              : :   ||:         |||        |  ::   :: :  :|:     ||    |: |
Db       1132 VSQLGAAGE-------VFNFGDFGDEDDHAASFGGLAAA-AAGAAGVARKRAFHGDDPFG 1183

Qy       1189 LSGIPIKHGNITMEM 1203
                | | |  ::|::|
Db       1184 -EGPPEKK-DLTLDM 1196


RESULT 6
Q9PZ52
ID   Q9PZ52      PRELIMINARY;      PRT;  1190 AA.
AC   Q9PZ52;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Homolog of HSV-1 glycoprotein.
OS   Turkey herpesvirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Marek's disease-like viruses.
OX   NCBI_TaxID=10390;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HPRS24;
RA   Kato K., Jang H., Izumiya Y., Cai J., Tsushima Y., Miyazawa T.,
RA   Kai C., Mikami T.;
RT   "Identification and Transcriptional Analysis of the Marek's Disease
RT   Virus Serotype 2 Genes Homologous to the Glycoprotein B (UL27), the
RT   ICP18.5 (UL28) and the Major DNA-binding Protein (UL29) Genes of
RT   Herpes Simplex Virus Type 1.";
RL   J. Vet. Med. Sci. 0:0-0(1999).
DR   EMBL; AB024711; BAA83753.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
SQ   SEQUENCE   1190 AA;  130339 MW;  D9291E2EEAA7F59D CRC64;

  Query Match            44.4%;  Score 2796.5;  DB 12;  Length 1190;
  Best Local Similarity  45.3%;  Pred. No. 2.6e-221;
  Matches  546;  Conservative  235;  Mismatches  396;  Indels   29;  Gaps   13;

Qy          6 KTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVV 65
              |:| :  ||:|||||     :  |::|  ||:| | : |:|||:  ||||  |  ::| |
```

```
Db          5 KSVKITGGPIGYVYAAPTSSMPAEDLSIFAAKSNDCEDAILPLVSGLTVEADFMWNVAAV 64

Qy         66 SGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRC 125

Db         65 AGTKTTGLGSGSTTLKLVPTHYHPCVFVFHGGECIKPCTKAPNLTKACDLARGRFGYS-T 123

Qy        126 QGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINH 185

Db        124 YSPPVATSFETTGEQICDNLGMNPQETMLYLVVTELFKEAVYLCNSYLHYGGTGVVSING 183

Qy        186 GDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPM 245

Db        184 VDVRRIPLYPLHLVFPDFNRVASDPFSTKPRALGEGALMPKAFYNDSLCRLLHGYVLSTA 243

Qy        246 AVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNST 305

Db        244 AVGLRVRNVDAIARCAAHLSFDENHEGTLLPADTAFNAFTPIESASKSQYKAGKKEGMEL 303

Qy        306 SKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRLNAL 365

Db        304 S----GGGYERRTASLMASDATLSIENVIATSVYEESIPDVKKWPLYCSPIGYTDRVEAL 359

Qy        366 GSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKD-GGPGPSFNRFYQFAGPHLAANP 424

Db        360 SAYMARVAGLVGAMVFSSNSVIYMTEVGEASSADGKDTSTTAASFYRFFQIAAPHLASNP 419

Qy        425 QTDRDGHVLS----SQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480

Db        420 LIDRDGKPIPGEDLSKATSASPSEYSLDYLILACGFCPQLLARFLFYLERCDGGSQACHH 479

Qy        481 G-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNS 539

Db        480 DLDTVKFVSSAMDADVPCDLCDKASRIYCAHTTIKRLEYRLPKFGYQMRGAMGLFGSMTN 539

Qy        540 QYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSS 599

Db        540 NYCDVNALGSYAPFSTLKR--SEGETSRSVMQDTYKLTVERVMKALEKEGLLTCEDPADM 597

Qy        600 EGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRTDYKIREGLSEATHSMALTFDP 659

Db        598 TPADAIIRDGKSFMRALSTMANIIESEAGQLMRNLTEIREYNIREGLGDATHTLSLAIEP 657

Qy        660 YSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFN 719

Db        658 YSSGICPVLSFLSRRTIIAVVQDMALSQCSMIMHGQQVEARNFRTQFQAVLRRRVLDLQN 717

Qy        720 GGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSG 779

Db        718 AGFITSKNITVTLEDQQISVPDPSKSQHDPLAVHMEGDLVKVTFEIFREFKVKNKVMFVG 777

Qy        780 NCTN-LSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNP 838

Db        778 GVSSTVSDATKSRLAGMIEAYQRPAKAMHVLNGPLGFALKRYHTQLFPNVKMPNGTTPNA 837

Qy        839 QWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLI 898

Db        838 LWFWILLQRNQLPAGILSKDAEENTSFIKRFTNSYADMNYINISPTCFGELAQFYLANTI 897

Qy        899 LKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPE 958

Db        898 LRYCSHKHFFINTISALVATSKRPRDPAMVLPWIEHILTQGSDVAPAAQQLLKNICDHKE 957

Qy        959 LWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFT 1018

Db        958 AWCAAFSSTNLVGPIMASKPFVVIAVSISKYHGMAGSTKVFQSGNWGNIMGGRNVCSIMS 1017

Qy       1019 FDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVG 1078

Db       1018 FNRTHRFVMACPRVGFVSEQTGFSSGLKETTLVDRARAVLSEECGAPHAAVYMIALKMVG 1077

Qy       1079 ARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDG 1138

Db       1078 DRVRQMELDDWMEITNDAYISSLIDELNKQ-VEGCEGGWSVDAAAILAKEMVDMSKALPL 1136

Qy       1139 ETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHG- 1197

Db       1137 DGPT---FNYDALDENTERRAD-----GPSI---LEPTLKRPCSD--VFDLEPVPEKRAP 1183

Qy       1198 NITMEM 1203

Db       1184 GLSVDM 1189


RESULT 7
Q9QTB9
ID   Q9QTB9      PRELIMINARY;      PRT;  1190 AA.
AC   Q9QTB9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UL29 protein.
GN   ORF 37 OR UL29.
OS   Marek's disease virus serotype 2 MDV2, and
OS   Gallid herpesvirus 3.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Marek's disease-like viruses.
OX   NCBI_TaxID=36353, 35250;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Gallid herpesvirus 1 (serotype 2); STRAIN=HPRS24;
RA   Jang H., Cai J., Izumiya Y., Murakami Y., Mochizuki M., Song C.,
RA   Lee Y., Kai C., Takahashi E., Mikami T.;
RT   "The complete DNA sequence and transcription map of the unique long
RT   genome region of Marek's disease virus type 2.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Gallid herpesvirus 3; STRAIN=HPRS24;
RA   Izumiya Y., Jang H., Ono M., Mikami T.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Gallid herpesvirus 3; STRAIN=HPRS24;
RA   Izumiya Y., Jang H., Ono M., Mikami T.;
RT   "A Complete Genomic DNA Sequence of Marek's Disease Virus Type 2,
RT   Strain HPRS24.";
RL   Curr. Top. Microbiol. Immunol. 0:0-0(2000).
DR   EMBL; AB024414; BAA82925.1; -.
DR   EMBL; AB049735; BAB16539.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
SQ   SEQUENCE   1190 AA;  130357 MW;  05589C36CA20BDE9 CRC64;

  Query Match           44.4%;  Score 2793.5;  DB 12;  Length 1190;
  Best Local Similarity   45.2%;  Pred. No. 4.5e-221;
  Matches  545;  Conservative  236;  Mismatches  396;  Indels   29;  Gaps   13;

Qy          6 KTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVV 65

Db          5 KSVKITGGPIGYVYAAPTSSMPAEDLSIFAAKSNDCEDAILPLVSGLTVEADFMWNVAAV 64

Qy         66 SGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRC 125

Db         65 AGTKTTGLGSGSTTLKLVPTHYHPCVFVFHGGECIKPCTKAPNLTKACDLARGRFGYS-T 123

Qy        126 QGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINH 185

Db        124 YSPPVATSFETTGEQICDNLGMNPQETMLYLVVTELFKEAVYLCNSYLHYGGTGVVSING 183

Qy        186 GDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPM 245

Db        184 VDVRRMPLYPLHLVFPDFNRVASDPFSTKPRALGEGALMPKAFYNDSLCRLLHGYVLSTA 243

Qy        246 AVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNST 305

Db        244 AVGLRVRNVDAIARCAAHLSFDENHEGTLLPADTAFNAFTPIESASKSQYKAGKKEGMEL 303
```

```
y       306 SKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRLNAL 365
b       304 S----GGGYERRTASLMASDATLSIENVIATSVYEESIPDVKKWPLYCSPIGYTDRVEAL 359

y       366 GSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKD-GGPGPSFNRFYQFAGPHLAANP 424
b       360 SAYMARVAGLVGAMVFSSNSVIYMTEVGEASSADGKDTSTTAASFYRFFQIAAPHLASNP 419

y       425 QTDRDGHVLS----SQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480
b       420 LIDRDGKPIPGEDLSKATSASPSEYSLDYLILACGFCPQLLARFLFYLERCDGGSQACHH 479

y       481 G-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNS 539
b       480 DLDTVKFVSSAMDADVPCDLCDKASRIYCAHTTIKRLEYRLPKFGYQMRGAMGLFGSMTN 539

y       540 QYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSS 599
b       540 NYCDVNALGSYAPFSTLKR--SEGETSRSVMQDTYKLTVERVMKALEKEGLLTCEDPADM 597

y       600 EGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDP 659
b       598 TPADAIIRDGKSFMRALSTMANIIESEAGQLMRNLTEIREYNIREGLGDATHTLSLAIEP 657

y       660 YSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFN 719
b       658 YSSGICPVLSFLSRRTIAVVQDMALSQCSMIMHGQQVEARNFRTQFQAVLRRRVLDLQN 717

y       720 GGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSG 779
b       718 AGFITSKNITVTLEDQQISVPDPSKSQHDPLAVHMEGDLVKVTFEIFREFKVKNKVMFVG 777

y       780 NCTN-LSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNP 838
b       778 GVSSTVSDATKSRLAGMIEAYQRPAKAMHVLNGPLGFALKRYHTQLFPNVKMPNGTTPNA 837

y       839 QWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLI 898
b       838 LWFWILLQRNQLPAGILSKDAEENTSFIKRFTNSYADMNYINISPTCFGELAQFYLANTI 897

y       899 LKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPE 958
b       898 LRYCSHKHFFINTISALVATSKRPRDPAMVLPWIEHILTQGSDVAPAAQQLLKNICDHKE 957

y       959 LWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFT 1018
b       958 AWCAAFSSTNLVGPIMASKPFVVIAVSISKYHGMAGSTKVFQSGNWGNIMGGRNVCSIMS 1017

y      1019 FDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVG 1078
b      1018 FNRTHRFVMACPRVGFVSEQTGFSSGLKETTLVDRARAVLSEECGAPHAAVYMIALKMVG 1077

y      1079 ARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDG 1138
b      1078 DRVRQMELDDWMEITNDAYISSLIDELNKQ-VEGCEGGWSVDAAAILAKEMVDMSKALPL 1136

y      1139 ETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHG- 1197
b      1137 DGPT---FNYDALDENTERRAD-----GPSI---LEPTLKRPCSD--VFDLEPVPEKRAP 1183

y      1198 NITMEM 1203
b      1184 GLSVDM 1189


ESULT 8
9E6P0
D   Q9E6P0      PRELIMINARY;      PRT;  1191 AA.
C   Q9E6P0;
T   01-MAR-2001 (TrEMBLrel. 16, Created)
T   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
T   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  UL29 single stranded DNA binding protein-like protein.
GN  MDV042.
OS  Turkey herpesvirus.
OC  Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC  Alphaherpesvirinae; Marek's disease-like viruses.
OX  NCBI_TaxID=10390;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Md5;
RX  MEDLINE=20392152; PubMed=10933706;
RA  Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
RT  "The genome of a very virulent Marek's disease virus.";
RL  J. Virol. 74:7980-7988(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Md5;
RA  Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
RL  Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF243438; AAG14222.1; -.
DR  InterPro; IPR000635; Viral_DNA_bind.
DR  Pfam; PF00747; viral_DNA_bp; 1.
SQ  SEQUENCE   1191 AA;  130714 MW;  9AD425780FB07711 CRC64;

  Query Match           44.3%;  Score 2788;  DB 12;  Length 1191;
  Best Local Similarity 45.1%;  Pred. No. 1.3e-220;
  Matches  542;  Conservative  237;  Mismatches  395;  Indels   28;  Gaps   12;

Qy        1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
Db        1 MDGVGKSVKLCGGPIGYIYATPKCSVPVDELAILAAKSNDCDDAVLPLVAGLTVESDFVW 60

Qy       61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
Db       61 NVAAVAGTKTTGLGSGGTTLKLVPTHYHPCVFVFYGGDCIKPCTKAPNLTKACDLARERF 120

Qy      121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
Db      121 GYS-AYSSPAPTAFETTGQQICEALEMDAQNVMLYLVVTELFKEVIYLCNSFLHFGGSDV 179

Qy      181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
Db      180 VTINNADVRRIPIYPLHLVLPDFNRITNEPFSEKPRALGEGAIMPKAFYNDSLCRLLHGY 239

Qy      241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300
Db      240 VLSTTAVGLRVRNIDAIARGAAHLCFDENHEGTLLPADTTFTAFTPAAE---TTKGQSKM 296

Qy      301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLP 360
Db      297 GKREGSDVS-GGGYERRTASLMASDATLAIENVISASVYEDPIPDVNKWPIYCNPVGYAD 355

Qy      361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKD-GGPGPSFNRFYQFAGPH 419
Db      356 RIEALSAYMARVAGLVGAMVFSSNSVIYMTEVGEAGSTEGKETSTTAPSFYRFFQIAAPH 415

Qy      420 LAANPQTDRDGHVLS----SQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGA 475
Db      416 LSANPLVDRDGKPVSGENLSKSTSASQSEYSLDYLILACGFCPQLLARFLFYLERCDGGA 475

Qy      476 FTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVF 534
Db      476 KACHHDLDTVKFVSSAIDADMPCELCDKTSRIYCAHTTIKRLVYRLPKFGYQMRGAMGLF 535

Qy      535 GTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRG 594
Db      536 GAMTNNYCDVNALGSYAQFSTLKR--SEGEASRSVMQDTYRLTVERMMKALEKEGLLTCD 593

Qy      595 APCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMA 654
Db      594 DPTNMASADANIRDGNSFIRAISTMKNIIESEASQLMRNLTEIREYNIREGLGDANHTLS 653

Qy      655 LTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRF 714
```



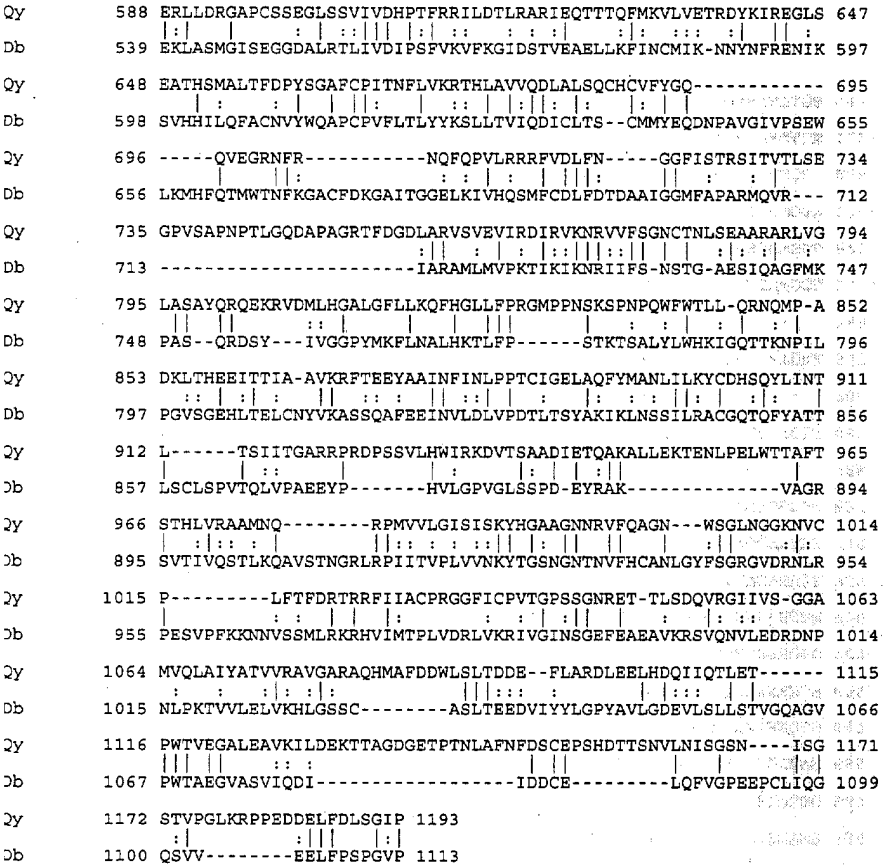
```
RESULT 9
Q9DPQ8
ID   Q9DPQ8      PRELIMINARY;      PRT;  1190 AA.
AC   Q9DPQ8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UL29 single stranded DNA binding protein.
GN   HVT037.
OS   Meleagrid herpesvirus 1 (herpesvirus of turkeys).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Marek's disease-like viruses.
OX   NCBI_TaxID=37108;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FC126;
RX   MEDLINE=20578232; PubMed=11134310;
RA   Afonso C.L., Tulman E.R., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
RT   "The genome of turkey herpesvirus.";
RL   J. Virol. 75:971-978(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FC126;
RA   Afonso C.L., Tulman E.R., Lu Z., Zsak L., Rock D.L., Kutish G.K.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF291866; AAG45767.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
SQ   SEQUENCE   1190 AA;  130054 MW;  B450CAF5FFF0D928 CRC64;

  Query Match           44.2%;  Score 2781.5;  DB 12;  Length 1190;
  Best Local Similarity  44.5%;  Pred. No. 4.4e-220;
  Matches  538;  Conservative  245;  Mismatches  391;  Indels   35;  Gaps   14;
```

```
                ::  ||  |  :  |  :||: :|  |::::  : :|::|  |: |  |::: |:   |
Db       1071 LALKMVGERVRQMELEDWIEITGDDYISSLIADLNEQAEQS-EGGWSIDVAMGLAK---- 1125

Qy       1132 KTTAGDGETPTN-LAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLS 1190
              :|  |      |:|   ||::  :   :  |:     | ::|     | |||  :  :|||
Db       1126 ETAAASAFIPSNGPTFNYEEWDDGVEKRSD-----GHSLSEG---GTKRP--NLSVFDLD 1175

Qy       1191 GIPIKHGNI 1199
                :|  |  :
Db       1176 PVPEKRAAV 1184


RESULT 10
Q9E1G3
ID   Q9E1G3      PRELIMINARY;      PRT;  1190 AA.
AC   Q9E1G3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UL29 major DNA binding protein.
GN   UL29.
OS   Meleagrid herpesvirus 1 (herpesvirus of turkeys).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Marek's disease-like viruses.
OX   NCBI_TaxID=37108;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FC126;
RX   MEDLINE=21195611; PubMed=11297687;
RA   Kingham B.F., Zelnik V., Kopacek J., Majerciak V., Ney E.,
RA   Schmidt C.J.;
RT   "The genome of herpesvirus of turkeys: comparative analysis with
RT   Marek's disease viruses.";
RL   J. Gen. Virol. 82:1123-1135(2001).
DR   EMBL; AF282130; AAG30069.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
SQ   SEQUENCE   1190 AA;  129952 MW;  A4A19239FB47DD82 CRC64;

 Query Match           43.9%;  Score 2764.5;  DB 12;  Length 1190;
 Best Local Similarity  44.3%;  Pred. No. 1.1e-218;
 Matches  536;  Conservative  245;  Mismatches  393;  Indels   35;  Gaps   14;

y          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
             |:   |:| :  ||:||||:    : : ::|:|  ||:  ||: ::|||:  ||||  |
b          1 MDGVGKSVKLVGGPIGYVYSMSKDAVPVDELSTFAAKCIDSEESILPLIAGLTVESEFVR 60

y         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
             ::| |:| :||||   ||||||  :|::|:||||:||  : | | |||||:||  |||||
b         61 NVAAVAGTKTTGLVNGGITLKLVPTHYHPNVFVFYGGDCIKPCSKAPNLTKACEYARERF 120

y        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
             |:      |   |:|||:| :|| :|||: : |:|||||| |||||:::|| ||:||| |
b        121 GYCPYSAP---GSVETSGKQICEQLGLDAQKTMLYLVVTELFKEAIYLCNSFLYYGGSDS 177

y        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
             |||  :| ||||:|: | :|| ||:  :||:   |::|:| | |  |:|  || |::
b        178 VTINDAEVRRIPLYPLHLVLPDFNRVSNEPFSERPRALGDGAVMPEAFFNDALCRLLYGY 237

y        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYF--QSSSSGTTTARGAR 298
             |:   || |||||: |||||||||| |||||||  :|| | |:| |   : : |    :  :
b        238 VLGTDAVGLRVRNIDAVARGAAHLCFDENHEGILLPADTTFTAFTPTADTGGKGQPKAGK 297

y        299 RNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGT 358
             |:        :  ||:|||  ||:||:|  |  | :  : :|||   |:|:||:|     |
b        298 RDGAE-----TQGGGYERRTASLMASDATLAIENVIASSVYEEPIPDVKKWPIFCNPVGY 352

y        359 LPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGG-PGPSFNRFYQFAG 417
               |: || :|  ||||::|||||| || :|:||| ::  ::  |: | | ||| ||:|:|
b        353 TDRVEALSAYMGRVAGLVGAMVFSSNSVIYMTEVGEASSSDGKESGVPAPSFYRFFQIAA 412

y        418 PHLAANPQTDRDGHVLSSQS----TGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDA 473
             |||||||  ||||  :|  :      |:::|:|:||| | |||   |||| ||||||||

Db       413 PHLAANPLVDRDGKPISGDALPKLAPSTSSEYSLDYLILACGFCPQLLARFLFYLERCDG 472

Qy       474 GAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIG 532
             ||     | :| |:  |:::|| ||:| :| ||||| ||  |:|:||  | :|
Db       473 GANACHHDLDTVKLVSSAIDADMPCELCDKTSRVYCAHTTVKRLEYRLPKFGYQMRGAMG 532

Qy       533 VFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLD 592
             :|| | : | | : ||:||:  |::  : ||::: |||||| |:||:   ||:| ||
Db       533 LFGAMTNNYCDVNALGSYAQFSTLKR--SEGEASRSVMQDTYRLTVERVMKALEKEGLLA 590

Qy       593 RGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHS 652
               | :     : | |  :|  : |:|  ||   :| |: |  |:  ||||| :| |:
Db       591 CDDPMNMAPADASIKDGASFIHAVSTMRNIIEGEASQLMRNLTEIRESNIREGLCDANHT 650

Qy       653 MALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRR 712
             ::|| :||| ||| :|| :|| :|||||||  :|||| ||| |||  |||| |||
Db       651 LSLTIEPYSSGFCPVLSFLARRTTIAVIQDMALSQCSMLMHGQQVEARNFRTQFQSVLRR 710

Qy       713 RFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
             | ::| | |||::: || || :  |: :|:  |  |      :||| ||::|:|::||
Db       711 RVLELQNAGFITSKDITATLEDQHVAVPDPSRSQYDPTVINMEGDLTRVTIEIIRELKVK 770

Qy       773 NRVVFSGNCT-NLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPP 831
             |||:| |      |:||::|| :  |||| | ::::|: ||| :|::|  ||     |
Db       771 NRVIFGGGVIGTASDAAKSRLANMIEAYQRPTKTMNVLNGPLGFAIKRYHSALFPNVKMP 830

Qy       832 NSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQ 891
             |  || |||||||||||  :|:|  :|:||  || :|:|| |  |:|||  |||
Db       831 NGAVPNAHWFWTLLQRNQLPDTMLSKEDQEKALFIKKFTNVYADMNYINISPTGFGDLAQ 890

Qy       892 FYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLE 951
             |::||:|||| |  : |||:::::  :::|||| :|| || | :|   |:  |:|:
Db       891 FFLANTILKYCSHKHFFINTISALVAASKKPRDPAIVLPWIDKTITQGRDVSLAAQQLIG 950

Qy       952 KTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGK 1011
               :  |  |::|:|| : |: :::|:|:|||||| ||: :|||:||| : ||:
Db       951 TIADRKDAWCATFSATNLVGSVMSTKPLIVIGVSISKYHGMAGSTKVFQSGNWGNIMGGR 1010

Qy      1012 NVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYA 1071
             ||| | :|||| |::|||| |||:  || ||| :| || : | |:        :|
Db      1011 NVCSLMSFDRTHRYVMACPRVGFVAEHTGFSSGLKEATLMDRARAILSEEGGAPHAEVYM 1070

Qy      1072 TVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDE 1131
               :: || | :  | :|: : |::::  : :|:|  |: |  |::: |:   |
Db      1071 LALKMVGERVRQMELEDWIEITGDDYISSLIADLNEQAEQS-EGGWSIDVAMGLAK---- 1125

Qy      1132 KTTAGDGETPTN-LAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLS 1190
             :| |      |:|   ||::  :   :  |:     | ::|     | |||  :  :|||
Db      1126 ETAAASAFIPSNGPTFNYEEWDDGVEKRSD-----GHSLSEG---GTKRP--NLSVFDLD 1175

Qy      1191 GIPIKHGNI 1199
                :|  |  :
Db      1176 PVPEKRAAV 1184


RESULT 11
Q9IBU6
ID   Q9IBU6      PRELIMINARY;      PRT;  1191 AA.
AC   Q9IBU6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UL29.
GN   UL29.
OS   Turkey herpesvirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Marek's disease-like viruses.
OX   NCBI_TaxID=10390;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GA;
RX   MEDLINE=92237304; PubMed=1315048;
RA   Jones D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;
```

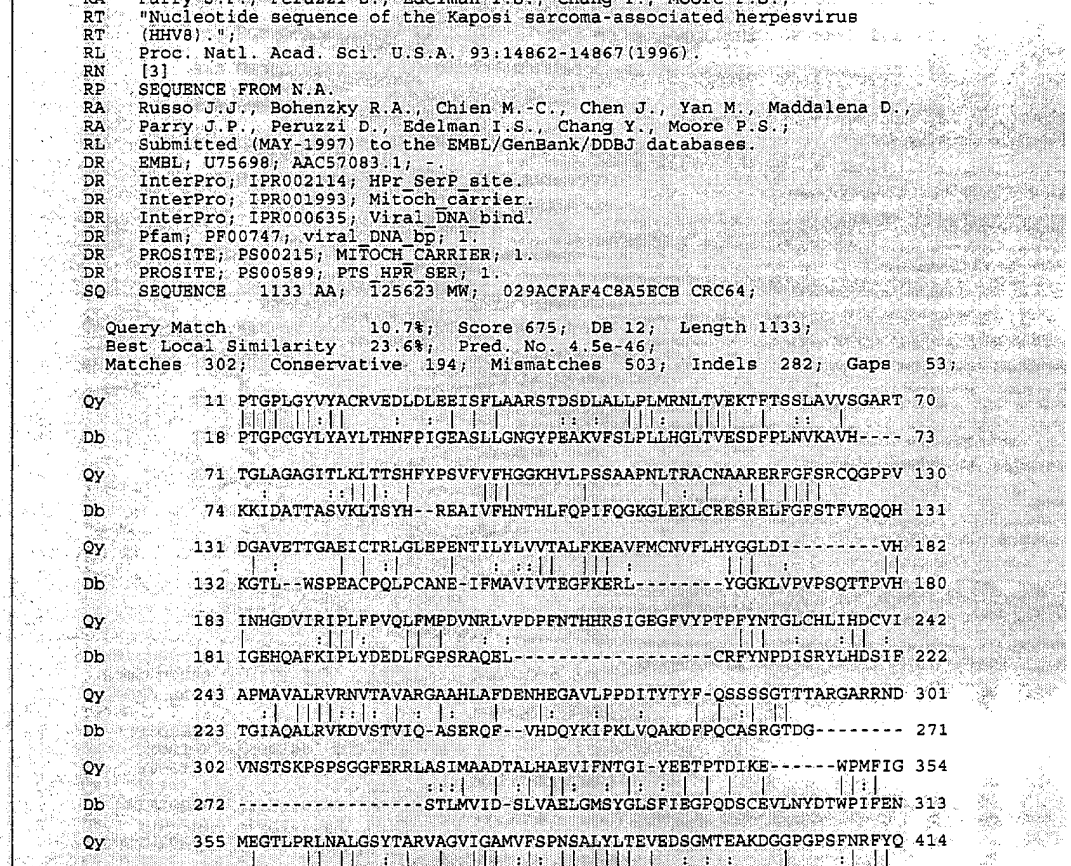
```
RT   "Marek disease virus encodes a basic-leucine zipper gene resembling
RT   the fos/jun oncogenes that is highly expressed in lymphoblastoid
RT   tumors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GA;
RX   MEDLINE=20283955; PubMed=10823954;
RA   Lee L.F., Wu P., Sui D., Ren D., Kamil J., Kung H.J., Witter R.L.;
RT   "The complete unique long sequence and the overall genomic
RT   organization of the GA strain of Marek's disease virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:6091-6096(2000).
DR   EMBL; AF147806; AAF66764.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
SQ   SEQUENCE   1191 AA;  130946 MW;  AC04A0158B0797A3 CRC64;

  Query Match           43.1%;  Score 2710;  DB 12;  Length 1191;
  Best Local Similarity   44.1%;  Pred. No. 3.6e-214;
  Matches  535;  Conservative  236;  Mismatches  400;  Indels   42;  Gaps   16;

Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISF------LAARSTDSDLALLPLMRNLTV 54
              |:   |:| :  ||:  :|  |         ::|      |||:| | | |:|||:  |||
Db          1 MDGVGKSVKLCGGPIA-IYMRR------PNVAFPWMNCDLAAKSNDCDDAVLPLVAGLTV 53

Qy         55 EKTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACN 114
             |  |  ::| |:| :|||   | |||  :|::| ||||:||  : | : |||||:|||:
Db         54 ESDFVWNVAAVAGTKTTGLGSGGTTLKLVPTHYHPCVFVFYGGDCIKPCTKAPNLTKACD 113

Qy        115 AARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLH 174
              ||||||:|    |   | |||| :||  | :: :| :|||||| |||| :::|| |||
Db        114 LARERFGYS-AYSSPAPTAFETTGQQICEALEMDAQNVMLYLVVTELFKEVIYLCNSFLH 172

Qy        175 YGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLC 234
             :|| |:| ||: || |||::|: | :|| ||:  :||:   |::||| : |  |||  ||
Db        173 FGGGSDVVTINNADVRRIPIYPLHLVLPDFNRITNEPFSEKPRALGEGAIMPKAFYNDSLC 232

Qy        235 HLIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTA 294
              |:|  |::  || |||||: |:|||||| |||||||  :|| | |:|  |  ::     |
Db        233 RLLHGYVLSTTAVGLRVRNIDAIARGAAHLCFDENHEGTLLPADTTFTAFTPAAE---TT 289

Qy        295 RGARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIG 354
             :|  :      |  |  ||:||| ||:||:|  |  | : :  :||:   |: :||::
Db        290 KGQSKMGKREGSDVS-GGGYERRTASLMASDATLAIENVISASVYEDPIPDVNKWPIYCN 348

Qy        355 MEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKD-GGPGPSFNRFY 413
              |   |: || :| |||||::|||||| || :|:||| ::| ||  :     ||| |||:
Db        349 PVGYADRIEALSAYMARVAGLVGAMVFSSNSVIYMTEVGEAGSTEGTETSTTAPSFYRFF 408

Qy        414 QFAGPHLAANPQTDRDGHVLS----SQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLE 469
              | |||:|||  ||||  :|     |:|| :| :|:|:||| | |||   |||| |||||
Db        409 QIAAPHLSANPLVDRDGKPVSGENLSKSTSASQSEYSLDYLILACGFCPQLLARFLFYLE 468

Qy        470 RCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATR 528
              |||  |     |  | :|:|:   |:::|| ||:| :|  |||||: ||  |:|:||   |
Db        469 RCDGGAKASHHDLDTVKFVSSAIDADMPCELCDKTSRIYCAHTTIKRLVYRLPKFGYQMR 528

Qy        529 QPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQE 588
              :|:|| | :  | | : |:|| :  |::   : ||::: ||||| |:||:   |||
Db        529 GAMGLFGAMTNNYCDVNALVSYAQFSTLKR--SEGEASRSVMQDTYRLTVERMMKALEKE 586

Qy        589 RLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGL-S 647
              ||    | :     : | |  :| | : |::   ||    :| |: | | |:| | |:  |
Db        587 GLLTCDDPTNMASADANIRDGNSFIRAISTMKNIIESEASQLMRNLTEIREYNITRGVWG 646

Qy        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707
             :| |:::|  :||:   ||:  || :|| :|||||:|||||  |  ||||| |||| |||
Db        647 DANHTLSLAVEPYASGICPVLAFLSRRTIIAVVQDMALSQCSIVMQGQQVEARNFRTQFQ 706

Qy        708 PVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIR 767
              ||:|| ::| | |||:::|||||  :  :   |:|:  |          :||| :|:|| |
Db        707 AVLKRRVLELQNAGFITSKTITVTLEDQQICVPDPSKSQYDSVISNMEGDLVKVTVEIFR 766

Qy        768 DIRVKNRVVFSGNCTN-LSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826
             ::|||:|| |      ||| ::|| |:  |||| |  : :|:|  ||| :|:: ||||
Db        767 ELKVKNKVLFGGGIAGAASEATKSRLAGMVEAYQRPTKTMHVLNGPLGFAVKRYHTLLFP 826

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886
                ||  :|| ||| || |||:|| |: ||      :|:||: || :|:|: |||
Db        827 DVKMPNGATPNALWFWILLLRNQLPAGILSKEEEDKSLFIKKFTKSYADMNYINISPTCF 886

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
              |:|||||:|| ||||| |  : |||:::::  :||||||: || || : :|   |:   |
Db        887 GDLAQFYLANTILKYCSHKHFFINTISALVAASRRPRDPAIVLPWIERPITKGQDVAPAA 946

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
              : |:   :  ::|   |||:|| : |  :|||:|||||||||| || : |||:|||:|||
Db        947 QQLIASMSDHKDIWCATFSSTNLVGSIMTTKPFVVIGISISKYHGMAGSTKVFQSGNWGN 1006

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
              : ||:||| | :|||| |::: ||| |||:        ||| :|||| |:|| ::   |
Db       1007 IMGGRNVCSLMSFDRTHRYVMTCPRVGFVAEQPIFSSGIKETTLIDRVRMVLSEESAAPH 1066

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAV 1126
               :|   :: || | :  | :||: :|:|:::  ::||: | ::  |  |  : :|:
Db       1067 AAVYMLALKMVGDRVRQMELEDWMEITNDEYISSLIDELNKQ-VEEAEGGWNADAAMTLA 1125

Qy       1127 KILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDEL 1186
                |  :    :  :|  |:||:|: : :  ::   || :|      ||||  :  :
Db       1126 KEMVNMAMSIPTDGPT---FDFDACDENLEGHADGQTISETN--------LKRP--NMNV 1172

Qy       1187 FDLSGIPIKHGNI 1199
              |||  || |  :
Db       1173 FDLEPIPEKRAPV 1185


RESULT 12
Q9QH63
ID   Q9QH63      PRELIMINARY;      PRT;   999 AA.
AC   Q9QH63;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   DNA binding protein.
OS   Gallid herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Infectious laryngotracheitis-like viruses.
OX   NCBI_TaxID=10386;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SA-2;
RA   Johnson M.A.;
RT   "Sequence of the infectious laryngotracheitis virus (SA-2 strain)
RT   unique long region UL28 to UL43.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF168792; AAD56201.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
SQ   SEQUENCE   999 AA;  108977 MW;  089FAED1C8BD674B CRC64;

  Query Match           23.0%;  Score 1445;  DB 12;  Length 999;
  Best Local Similarity   30.5%;  Pred. No. 8.7e-110;
  Matches  351;  Conservative  177;  Mismatches  376;  Indels  246;  Gaps   18;

Qy         13 GPLGYVYACRVEDLDLEEISFLAARSTDS-DLALLPLMRNLTVEKTFTSSLAVVSGARTT 71
              ||:||||    |:: | :|   | |:| |   |: ||:| ||||  |  ::| | | :::
Db         22 GPIGYVYARPREEIDGKEWKILCAKSQDQPSCAIAPLIRGLTVESDFKPNVAAVIGTKSS 81

Qy         72 GLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRC-QGPPV 130
              |: |    |:  ||  :|::|||||: : |:|  |||:  |  |||||||       ||
Db         82 GVVGGNCTAILSPCHFSTTVYIFHGGECIPPTSFTPNLTKICEDARERFGFSSLPPNGPV 141

Qy        131 DGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINHGDVIR 190
               :|||| ||| | ::|: |:|||||  | |||:|| ::||| :: ::||    | |
Db        142 PNAIETTGEEICKSLNMDPDKTMLYLVVAEPFCEAVYVCNTYIHYGSVETLYINSKHVTR 201
```

```
Qy        191 IPLFPVQLFMPDVN-RLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVAL 249
Db        202 IPIYPVQMYMPDIALRICRNPFDSNSRNIGEGCIYPTPLFNKALNRLIHGATLGTQGQSL 261

Qy        250 RVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPS 309
Db        262 RTRDLEAVARGAAMLAFDGSFEGCVLTSDKTFTEL-------------AQNIQQTGPPK 307

Qy        310 PSGGFERRLASIMAADTALHAEVIFNTGIYEETPTD------IKEWPMFIGMEGTLPRLN 363
Db        308 QNVEVERRAACSLAAELALATRLSVSCAPY---PFDNGCGLPYEEWPLFCDCKTQMERVQ 364

Qy        364 ALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPG----PSFNRFYQFAGPH 419
Db        365 ALDRFSAELAGIVGAAIFSNNSPLYASEVVDGGAPE-----PGERASSGLTRFYLTCGLH 419

Qy        420 LAANPQTDRDGH-----VLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAG 474
Db        420 LLGCPQVDYSGKRVWEGTCPSPLTPQNTFEYDPEHLAYACGFSPELLSRIIFYLEKCSKQ 479

Qy        475 AFTGGHGDALKYVTGTFDSEIPCS-------LCEKHTRPVCAHTTVHRLRQRMPRFGQAT 527
Db        480 LSTRPN---MNLISISNSGKAPCTGAVPECRWCNDESRNHCIRYTMQRLRTRFPIPRTMR 536

Qy        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
Db        537 RSPIAVIGAVDGDYTDCDILGNFAPYSQLKRAGD-GEPAKAVMNDTYRGIGWRVFQYLVS 595

Qy        588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647
Db        596 EGLINKDTGEDTHNISSL----AELKTVFEKIQNFVGSECSKFISALSGVRAYHYKEHLS 651

Qy        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707
Db        652 SIVHTFG-SPEPYSSAYCPVLSLLCAQTRSILFQDLILSQIHGTFDTRQPESKMFRSSAM 710

Qy        708 PVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIR 767
Db        711 PALRSAFNGMLDKGFLSGKYEPFSISVPCVNAPDTMRPNTEQAIAQYEYSLSRGQVLKLK 770

Qy        768 DIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPR 827
Db        771 EFKVKNRVVFNGSSTSNS---------------------------------------- 788

Qy        828 GMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIG 887
Db        789 ------------------------------------------------------------ 788

Qy        888 ELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAK 947
Db        789 ------------------------------------------------------------ 788

Qy        948 ALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGL 1007
Db        789 ----------------------CRVVLSSKPIVMLGIMISKYVGQQNSTTVFQSGNWGSL 826

Qy       1008 ---NGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAM 1064
Db        827 IGNSGSQSVNSALTNDPVRKFALACKRVGTILSSGQVASGFQENTIASQVRSLIDAGGS- 885

Qy       1065 VQLAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALE 1124
Db        886 -PTSIYTAVLRVLGEGMKDITAETWMAITDDKYLVNILVELREDISGS-QNGWSVATAQS 943

Qy       1125 AVKILDEKTT 1134
Db        944 MLSELEGKDT 953
```

```
ESULT 13
40913
D   O40913      PRELIMINARY;      PRT;  1132 AA.
C   O40913;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ORF 06.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97296220; PubMed=9151804;
RA   Neipel F., Albrecht J.C., Fleckenstein B.;
RT   "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
RT   human herpesvirus 8: determinants of its pathogenicity?";
RL   J. Virol. 71:4187-4192(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
RA   Friedman-Kien A.E., Fleckenstein B.;
RT   "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U93872; AAB62603.1; -.
DR   InterPro; IPR002114; HPr_SerP_site.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
DR   PROSITE; PS00215; MITOCH_CARRIER; 1.
DR   PROSITE; PS00589; PTS_HPR_SER; 1.
SQ   SEQUENCE   1132 AA;  125394 MW;  CF88E4F451919CBF CRC64;

  Query Match           10.9%;  Score 683.5;  DB 12;  Length 1132;
  Best Local Similarity 23.5%;  Pred. No. 9e-47;
  Matches  301;  Conservative  196;  Mismatches  500;  Indels  285;  Gaps   53;

Qy         11 PTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGART 70
Db         18 PTGPCGYLYAYLTHNFPIGEASLLGNGYPEAKVFSLPLLHGLTVESDFPPLNVKAVH---- 73

Qy         71 TGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPPV 130
Db         74 KKIDATTASVKLTSYH--REAIVFHNTHLFQPIFQGKGLEKLCRESRELFGFSTFVEQQH 131

Qy        131 DGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI--------VH 182
Db        132 KGTL--WSPEACPQLPCANE-IFMAVIVTEGFKERL--------YGGKLVPVPSQTTPVH 180

Qy        183 INHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVI 242
Db        181 IGEHQAFKIPLYDEDLFGPSRAQEL-----------------CRFYNPDISRYLHDSIF 222

Qy        243 APMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYF-QSSSSGTTTARGARRND 301
Db        223 TGIAQALRVKDVSTVIQ-ASERQF--VHDQYKIPKLVQAKDFPQCASRGTDG-------- 271

Qy        302 VNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGI-YEETPTDIKE------WPMFIG 354
Db        272 ----------------STLMVID-SLVAELGMSYGLSFIEGPQDSCEVLNYDTWPIFEN 313

Qy        355 MEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQ 414
Db        314 CETPDARLRALEVWHAEQALHIGAQLFAANSVLYLTRV--AKLPQKNQRGDANMYNSFYL 371

Qy        415 FAGPHLAANPQTDRDG-----HVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLE 469
Db        372 QHGLGYLSEATVKENGASAFKGVPVSALDGSS---YTLQHLAYASSFSPHLLARMCYYLQ 428

Qy        470 RCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPR-FGQAT 527
Db        429 FLPHHKNTNSQSYNVVDYV-GTAAPSQMCDLCQGQCPAVCINTLFYRMKDRFPPVLSNVK 487

Qy        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
Db        488 RDPYVITGTAGT-YNDLEILGNFATFRERFEEEGNPVEDAP---KYTYWQLCQNI-----T 538
```



```
RESULT 14
P88904
ID   P88904      PRELIMINARY;      PRT;  1133 AA.
AC   P88904;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ORF 6.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97094384; PubMed=8939871;
RA   Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
RT   "Molecular mimicry of human cytokine and cytokine response pathway
RT   genes by KSHV.";
RL   Science 274:1739-1744(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97121480; PubMed=8962146;
RA   Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT   "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
RT   (HHV8).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U75698; AAC57083.1; -.
DR   InterPro; IPR002114; HPr_SerP_site.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
DR   PROSITE; PS00215; MITOCH_CARRIER; 1.
DR   PROSITE; PS00589; PTS_HPR_SER; 1.
SQ   SEQUENCE   1133 AA;  125623 MW;  029ACFAF4C8A5ECB CRC64;

  Query Match           10.7%;  Score 675;  DB 12;  Length 1133;
  Best Local Similarity   23.6%;  Pred. No. 4.5e-46;
  Matches  302;  Conservative  194;  Mismatches  503;  Indels  282;  Gaps   53;

Qy        11 PTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGART 70
             |||| ||:||    :  : | | |      :: :  |||:  ||||  |  ::  |
Db        18 PTGPCGYLYAYLTHNPPIGEASLLGNGYPEAKVFSLPLLHGLTVESDFPLNVKAVH---- 73

Qy        71 TGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPPV 130
             :    ::|||: |      |||          |  : |  :|| ||||
Db        74 KKIDATTASVKLTSYH--REAIVFHNTHLFQPIFQGKGLEKLCRESRELFGFSTFVEQQH 131

Qy       131 DGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI--------VH 182
              | :    | | :|    |  : ::|| ||| :       |||  :        ||
Db       132 KGTL--WSPEACPQLPCANE-IFMAVIVTEGFKERL--------YGGKLVPVPSQTTPVH 180

Qy       183 INHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVI 242
             |      :|||:  || |  : :          |||  :    :||
Db       181 IGEHQAFKIPLYDEDLFGPSRAQEL----------------CRFYNPDISRYLHDSIF 222

Qy       243 APMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYF-QSSSSGTTTARGARRND 301
              :| |||||::|  : :|:    |  :  :|  :  | | :| ||
Db       223 TGIAQALRVKDVSTVIQ-ASERQF--VHDQYKIPKLVQAKDFPQCASRGTDG-------- 271

Qy       302 VNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGI-YEETPTDIKE------WPMFIG 354
                        :::|  | :| ||:  : |: : | |  |      ||:|
Db       272 ----------------STLMVID-SLVAELGMSYGLSFIEGPQDSCEVLNYDTWPIPEN 313

Qy       355 MEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQ 414
             |  || ||  :  |  ||| : :  || |||  :  :  :  |  :  :|  ||
Db       314 CETPDARLRALEVWHAEQALHIGAQLFAANSVLYLTRV--AKLPQKNQRGDANMYNSFYL 371

Qy       415 FAGPHLAANPQTDRDG-----HVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLE 469
             |  :        :|  |  |||    :: : ||  |  ||||: :||:
Db       372 QHGLGYLSEATVKENGASAFKGVPVSALDGSS---YTLQHLAYASSFSPHLLARMCYYLQ 428

Qy       470 RCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPR-FGQAT 527
             |      : : || ||  |  |||:      || :|  :|:: | |
Db       429 FLPHHKNTNSQSYNVVDYV-GTAAPSQMCDLCQGQCPAVCINTLFYRMKDRFPPVLSNVK 487

Qy       528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
             |  :  || :  :|  :  | |||:  :  :|:  |  |  :||  :  |  |
Db       488 RDPYVITGTAGT-YNDLEILGNFATFREREEEGNPVEDAP---KYTYWQLCQNI-----T 538

Qy       588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647
             |:|  |      :  | :|||  |:|  ::   : : :|    :|:  :::  :|  || :
Db       539 EKLASMGISEGGDALRTLIVDIPSFVKVFKGIDSTVEAELLKFINCMIK-NNYNFRENIK 597

Qy       648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQ------------ 695
                 |  :   : |  | ||:    |  :: | :|:|: |:   |: | |
Db       598 SVHHILQFACNVYWQAPCPVFLTLYYKSLLTVIQDICLTS--CMMYEQDNPAVGIVPSEW 655

Qy       696 -----QVEGRNFR-----------NQFQPVLRRRFVDLFN-----GGFISTRSITVTLSE 734
                  |    ||:           : : | :  | |||:     ||  :   : |
Db       656 LKMHFQTMWTNFKGACFDKGAITGGELKIVHQSMFCDLFDTDAAIGGMFAPARMQVR--- 712

Qy       735 GPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVG 794
                              :||  : | : |::|||::||  | |  :|: :|  :
Db       713 ----------------------IARAMLMVPKTIKIKNRIIFS-NSTG-AESIQAGFMK 747

Qy       795 LASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLL-QRNQMP-A 852
              ||  ||      :: |     |   | |||      |   :  : |  : |  : |
Db       748 PAS--QRDSY---IVGGPYMKFLNALHKTLFP------STKTSALYLWHKIGQTTKNPIL 796

Qy       853 DKLTHEEITTIA-AVKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINT 911
               :: | :| :   ||  :: :  || ::| |  :   |:  : : ||: |  :|:   |
Db       797 PGVSGEHLTELCNYVKASSQAFEEINVLDLVPDTLTSYAKIKLNSSILRACGQTQFYATT 856

Qy       912 L------TSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFT 965
             |      | ::     |  | :     |: | | :||              |
Db       857 LSCLSPVTQLVPAEEYP-------HVLGPVGLSSPD-EYRAK-------------VAGR 894

Qy       966 STHLVRAAMNQ--------RPMVVLGISISKYHGAAGNNRVFQAGN---WSGLNGGKNVC 1014
             |  :|:: : |        ||:: : : ::|| |: ||  ||    |   :||    :|:
Db       895 SVTIVQSTLKQAVSTNGRLRPIITVPLVVNKYTGSNGNTNVFHCANLGYFSGRGVDRNLR 954

Qy      1015 P---------LFTFDRTRRFIIACPRGGFICPVTGPSSGNRET-TLSDQVRGIIVS-GGA 1063
             |         : :  | | |:       :  : | :||   |    :    |: ::
Db       955 PESVPFKKNNVSSMLRKRHVIMTPLVDRLVKRIVGINSGEFEAEAVKRSVQNVLEDRDNP 1014

Qy      1064 MVQLAIYATVVRAVGARAQHMAFDDWLSLTDDE--FLARDLEELHDQIIQTLET------ 1115
              :   :  :|: :|:           |||:::  :       | |:::  | |
Db      1015 NLPKTVVLELVKHLGSSC--------ASLTEEDVIYYLGPYAVLGDEVLSLLSTVGQAGV 1066

Qy      1116 PWTVEGALEAVKILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSN----ISG 1171
             |||  ||   :: :                | ||          |    |      |
Db      1067 PWTAEGVASVIQDI-----------------IDDCE---------LQFVGPEEPCLIQG 1099

Qy      1172 STVPGLKRPPEDDELFDLSGIP 1193
              :|         :|||    |:|
Db      1100 QSVV--------EELFPSPGVP 1113
```



```
Qy       735 GPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVG 794
Db       713 ----------------------IARAMLMVPKTIKIKNRIIFS-NSTG-AESIQAGFMK 747

Qy       795 LASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLL-QRNQMP-A 852
Db       748 PAS--QRDSY---IVGGPYMKFLNALHKTLFP------STKTSALYLWHKIGQTTKNPIL 796

Qy       853 DKLTHEEITTIA-AVKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINT 911
Db       797 PGVSGEHLTELCNYVKASSQAFEEINVLDLVPDTLTSYAKIKLNSSILRACGQTQFYATT 856

Qy       912 L------TSIIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFT 965
Db       857 LSCLSPVTQLVPAEEYP-------HVLGPVGLSSPD-EYRAK--------------VAGR 894

Qy       966 STHLVRAAMNQ--------RPMVVLGISISKYHGAAGNNRVFQAGN---WSGLNGGKNVC 1014
Db       895 SVTIVQSTLKQAVSTNGRLRPIITVPLVVNKYTGSNGNTNVFHCANLGYFSGRGVDRNLR 954

Qy      1015 P---------LFTFDRTRRFIIACPRGGFICPVTGPSSGNRET-TLSDQVRGIIVSGGAM 1064
Db       955 PESVPFKKNNVSSMLRKRHVIMTPLVDRLVKRIVGINSGEFEAEAVKRSVQNVLEDRD-- 1012

Qy      1065 VQLAIYATVVRAVGARAQHMAFDDWLSLTDDE--FLARDLEELHDQIIQTLET------P 1116
Db      1013 -NPNLPKTVVLEL---VKPPRWSSCASLTEEDVIYYLGPYAVLGDEVLSLLSTVGQAGVP 1068

Qy      1117 WTVEGALEAVKILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSN----ISGS 1172
Db      1069 WTAEGVASVIQDI------------------IDDCE---------LQFVGPEEPCLIQGQ 1101

Qy      1173 TVPGLKRPPEDDELFDLSGIP 1193
Db      1102 SVV--------EELFPSPGVP 1114
```

```
ESULT 15
Q992Z6
D   Q992Z6      PRELIMINARY;      PRT;  1103 AA.
C   Q992Z6;
T   01-JUN-2001 (TrEMBLrel. 17, Created)
T   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
T   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E   Single-stranded DNA-binding protein.
N   6.
S   Murid herpesvirus 4.
C   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C   Gammaherpesvirinae.
X   NCBI_TaxID=33708;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=MHV76;
X   MEDLINE=21232509; PubMed=11333912;
A   Macrae A.I., Dutia B.M., Milligan S., Brownstein D.G., Allen D.J.,
A   Mistrikova J., Davison A.J., Nash A.A., Stewart J.P.;
T   "Analysis of a novel strain of murine gammaherpesvirus reveals a
T   genomic locus important for acute pathogenesis.";
L   J. Virol. 75:5315-5327(2001).
N   [2]
P   SEQUENCE FROM N.A.
C   STRAIN=MHV76;
A   Davison A.J.;
L   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
R   EMBL; AF324455; AAK16699.1; -.
R   InterPro; IPR002114; HPr_SerP_site.
R   InterPro; IPR000531; TonB_boxC.
R   InterPro; IPR000635; Viral_DNA_bind.
R   Pfam; PF00747; viral_DNA_bp; 1.
R   PROSITE; PS00589; PTS_HPR_SER; 1.
R   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
W   DNA-binding.
```

```
FT   VARIANT     356    356       S -> G.
SQ   SEQUENCE   1103 AA;  123272 MW;  056359648CE27C03 CRC64;

  Query Match           10.4%;  Score 652.5;  DB 12;  Length 1103;
  Best Local Similarity 23.9%;  Pred. No. 3.1e-44;
  Matches  292;  Conservative  194;  Mismatches  535;  Indels  199;  Gaps   47;

Qy         1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
Db         1 METTNVCSQAPLGPAGYIYYQLVDDFPLEEASLLSTNFTSAKACLLPLVTGLTVEPGFNF 60

Qy        61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
Db        61 NVK----APCQKIMFESLTVKPT--KYFNHAIVFHNSNKIPPIFHGRGLEQLCEATRLMF 114

Qy       121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
Db       115 GFSSFKPDASKGQWKSS-----THLPQNPEKYIGWVVVAESFKERLICGELAMLSAPMKT 169

Qy       181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
Db       170 VTINNKLCFKVPLHDTRFLSSD--QLL------------------PFYDENVSRIMYES 208

Qy       241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300
Db       209 YYTNVAQAVRIRDVTGLIEALQERAVTEQYKTAKLASLKTY------------------ 249

Qy       301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETP-TDIKEWPMFIGMEGTL 359
Db       250 ----PAENCGAGNLEYAVIDAAVSELAVSHGLAFLEAPQEESPIQNYVQWPMF--SEDAT 303

Qy       360 P--RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSF-NRFYQFA 416
Db       304 PEGRIKALQEWNAKMAIHVHAQVLSANSVLYVTKVGQSA-GKAVGGKQDDMFLSRYYMQH 362

Qy       417 G-PHLAANPQTDRDGHVLSSQSTGSS---NTEFSVDYLALICGFGAPLLARLLFYLERCD 472
Db       363 GLGHL--NVQTRDENNTVVFPGVPDSCMNGSQYTVHHLAYAASMSPHILARLCYYLQMCQ 420

Qy       473 AG--AFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPR-FGQATRQ 529
Db       421 NQRLASTPQGTTVPAYVTTTANTSM-CDLCGGDLPAVCLHTLFFRLRDRFPQVLATQKRD 479

Qy       530 PIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQER 589
Db       480 PYVVTG-IAGPHNDMEILGNFGSF--KEKDDDQ--------ENGQRYSYWQLNANL-TER 527

Qy       590 LLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETR-DYKIREGLSE 648
Db       528 LEEIGVTASTPA-CQLITDVDSFLTVFKKIDAIVDEETIRFNDGLVRNNINYK--ESVKS 584

Qy       649 ATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQC------HCVFYGQQVEGRNF 702
Db       585 IAHVLQLQCNTNWLPPCPIIHNLFTRSFFTILQDMSFPLCVTHHTENPLNYG--TVSTWM 642

Qy       703 RNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVS 762
Db       643 SNHFQNLWSNFKSIWFDKGLLTCSDMRVVHTE--TSSDMENLAADLPKQRCAI-RLSRAQ 699

Qy       763 VEVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHG 822
Db       700 IFAPKALKIKNRIIFSNSSS--TESLSASFVKAGKAEN------PITAGPYMHFLTQLHK 751

Qy       823 LLFPRGMPPNSKSPNPQWFW-TLLQRNQMP------ADKLTHEEITTIAAVKRFTEEYAA 875
Db       752 QLFP------TTKTGAYYIWQTFMSSKKLPQAGSTNAKKLS-EFISYLMTSSLAHDE--- 801

Qy       876 INFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSI-------------ITGARR 921
Db       802 VNVLDYVPTTLMAYAKQRLNNAILRLCGQTQFYATTINFLQPTILSMPSLDYPHVAG--- 858

Qy       922 PRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVV 981
```

```
Db        859 ---PVPV-----QDLTTYMDMTRDQSATVIQSSCRED----------PAAVCKMRPIVT 899

Qy        982 LGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFDR----------------TRR 1024
              : : ::|| |: ||| :||:||  |   |: |      ||                ||
Db        900 IPMMVNKYSGSNGNNSIFQSGN-MGYFMGRGV------DRKLLPDIPRMRKHVNTSMRRR 952

Qy       1025 FIIACPRGGFICPVTGPSSGNRETTLSDQVR----GIIVSGGAMVQLAIYATVVRAVGAR 1080
              :  | |    :  :| |  |   |   :: |     || || |  :  :   :|: :|
Db        953 YAFASPMTESL--LTQPGKGTVRTYQLEKTRKIIHDIIASGNAEPE-NLVLELVKCLGHE 1009

Qy       1081 AQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL--EAVKILDEKTTAGDG 1138
                :  |       | ||   :   : : ::: |||      |   |:|: |   :: |
Db       1010 CATLTSD------DVEFYLGEFSFMVNDVMEVLETLKASGCAFSEESVQALLPTSSPGTS 1063

Qy       1139 ETPTNLAFNFDSCEPSHDTT 1158
              |:   :  |  :  |    |
Db       1064 ESFVEVTDNIITSLPEISVT 1083
```

Search completed: January 30, 2004, 13:14:03
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 30, 2004, 13:07:07 ; Search time 17 Seconds
(without alignments)
3327.831 Million cell updates/sec

Title: US-09-769-699-2
Perfect score: 6294
Sequence: 1 MENTQKTVTVPTGPLGYVYA.....DELFDLSGIPIKHGNITMEM 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6294	100.0	1204	1	DNBI_VZVD	P09246 varicella-z
2	3620.5	57.5	1209	1	DNBI_HSVEB	P28932 equine herp
3	3054	48.5	1196	1	DNBI_HSV1K	P17470 herpes simp
4	3049	48.4	1196	1	DNBI_HSV11	P04296 herpes simp
5	3047	48.4	1196	1	DNBI_HSV1F	P17469 herpes simp
6	3015.5	47.9	1196	1	DNBI_HSV2H	P89452 herpes simp
7	3013	47.9	1197	1	DNBI_HSV2	P36384 herpes simp
8	2933	46.6	1186	1	DNBI_HSVB2	P12639 bovine herp
9	1094.5	17.4	375	1	DNBI_HSVE1	Q03444 equine herp
10	669	10.6	1128	1	DNBI_EBV	P03227 epstein-bar
11	604	9.6	1128	1	DNBI_HSVSA	P24910 herpesvirus
12	492.5	7.8	1132	1	DNBI_HSV6U	P52338 human herpe
13	490.5	7.8	1132	1	DNBI_HSV6Z	P52538 human herpe
14	483	7.7	1131	1	DNBI_HSV7J	P52339 human herpe
15	479.5	7.6	1160	1	DNBI_SCMVC	P13215 simian cyto
16	477	7.6	1235	1	DNBI_HCMVA	P17147 human cytom
17	473	7.5	1193	1	DNBI_HSVT2	Q9wrl7 herpesvirus
18	412	6.5	1191	1	DNBI_MCMVS	P30672 murine cyto
19	121	1.9	1066	1	Z295_HUMAN	Q9ulj3 homo sapien
20	115	1.8	823	1	CEK3_CHICK	P18461 gallus gall
21	114	1.8	822	1	IF2C_EUGGR	Q9xek9 euglena gra
22	114	1.8	1941	1	YRM8_CAEEL	Q09417 caenorhabdi
23	113.5	1.8	1886	1	GP21_RAT	P11654 rattus norv
24	113	1.8	2505	1	FAS_RAT	P12785 rattus norv
25	110	1.7	3333	1	LMA3_MOUSE	Q61789 mus musculu
26	109	1.7	2845	1	APC_MOUSE	Q61315 mus musculu
27	108	1.7	827	1	XANP_XANS2	Q60106 xanthomonas
28	108	1.7	1902	1	P2P_LACPA	Q02470 lactobacill
29	107	1.7	1279	1	BCSC_PSEFL	P58937 pseudomonas
30	106.5	1.7	3298	1	PC16_HUMAN	Q96jq0 homo sapien
31	106.5	1.7	3567	1	ERY2_SACER	Q03132 saccharopol
32	106	1.7	773	1	PAC2_PSES3	P15558 pseudomonas
33	106	1.7	3149	1	TEGU_EBV	P03186 epstein-bar
34	105.5	1.7	1016	1	UVRA_DEIRA	Q46577 deinococcus
35	105.5	1.7	2114	1	MY9B_MOUSE	Q9qy06 mus musculu
36	105.5	1.7	4540	1	DYHC_PARTE	Q27171 paramecium
37	102.5	1.6	1902	1	P2P_LACLC	P15293 lactococcus
38	102	1.6	1038	1	V120_VZVD	P09277 varicella-z
39	102	1.6	1275	1	RFBC_MYXXA	Q50864 myxococcus
40	102	1.6	4590	1	FATH_HUMAN	Q14517 homo sapien
41	101.5	1.6	1478	1	BUD3_ASHGO	Q9hf61 ashbya goss
42	101.5	1.6	1902	1	P1P_LACLC	P16271 lactococcus
43	101	1.6	1229	1	CY43_TRYBB	Q99280 trypanosoma
44	101	1.6	1348	1	VGR2_COTJA	P52583 coturnix co
45	101	1.6	3649	1	ACVS_NOCLA	P27743 nocardia la

ALIGNMENTS

RESULT 1
DNBI_VZVD
ID DNBI_VZVD STANDARD; PRT; 1204 AA.
AC P09246;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein.
GN 29.
OS Varicella-zoster virus (strain Dumas) (VZV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10338;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86306657; PubMed=3018124;
RA Davison A.J., Scott J.E.;
RT "The complete DNA sequence of varicella-zoster virus.";
RL J. Gen. Virol. 67:1759-1816(1986).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X04370; CAA27912.1; -.
DR PIR; C27214; DNBE29.
DR InterPro; IPR000635; Viral_DNA_bind.
DR Pfam; PF00747; viral_DNA_bp; 1.
KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN_FING 497 510 C2HC-TYPE.
SQ SEQUENCE 1204 AA; 132139 MW; D2FEFE65DC0CC674 CRC64;

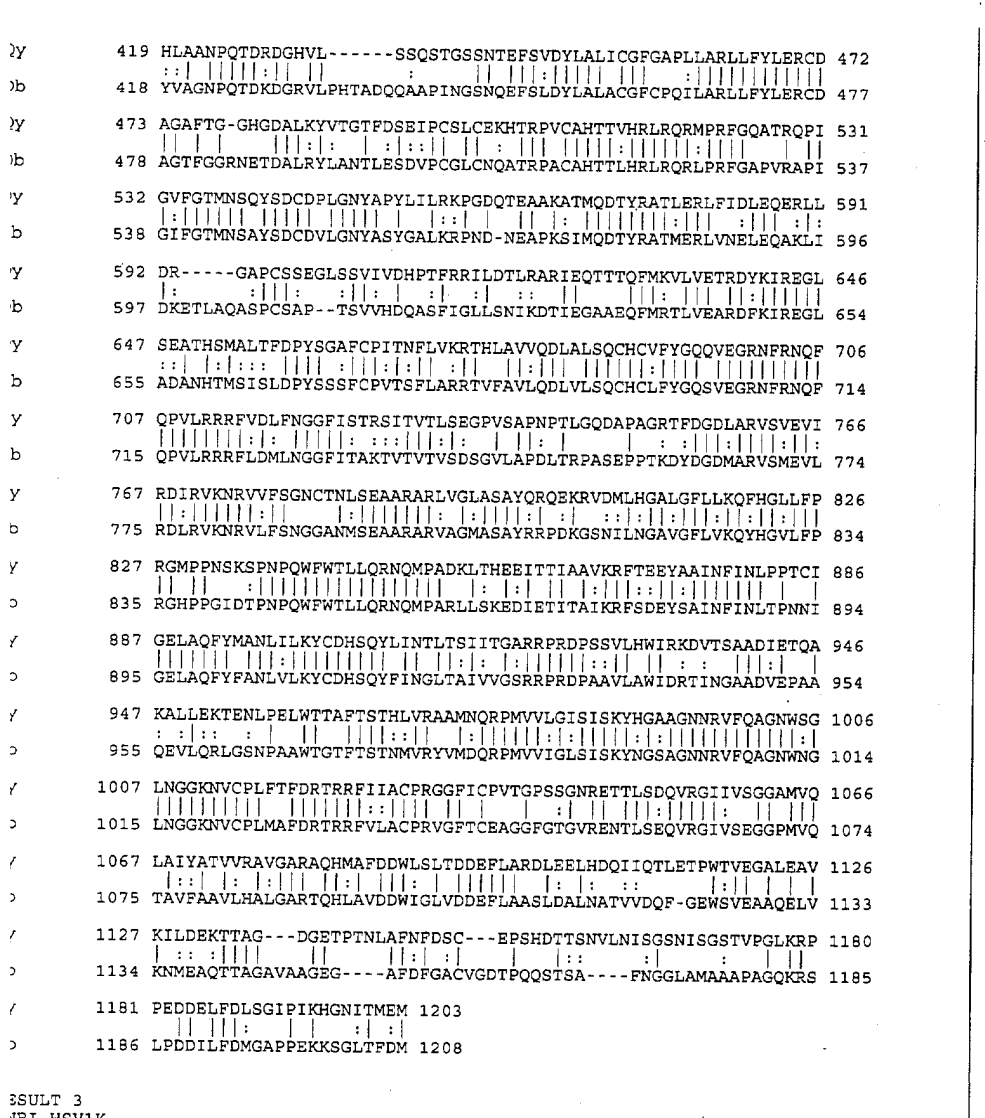
Query Match 100.0%; Score 6294; DB 1; Length 1204;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120

Qy        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
```

09/769,699 2a/1pg
Search Notes

```
Qy        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240

Qy        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300

Qy        301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLP 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLP 360

Qy        361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHL 420

Qy        421 AANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480

Qy        481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNSQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNSQ 540

Qy        541 YSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSE 600

Qy        601 GLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660

Qy        661 SGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFNG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFNG 720

Qy        721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGN 780

Qy        781 CTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW 840

Qy        841 FWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 FWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900

Qy        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960

Qy        961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFD 1020

Qy       1021 RTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGAR 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 RTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGAR 1080

Qy       1081 AQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGET 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGET 1140

Qy       1141 PTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHGNIT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 PTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHGNIT 1200

Qy       1201 MEM 1203
              |||
Db       1201 MEM 1203
```

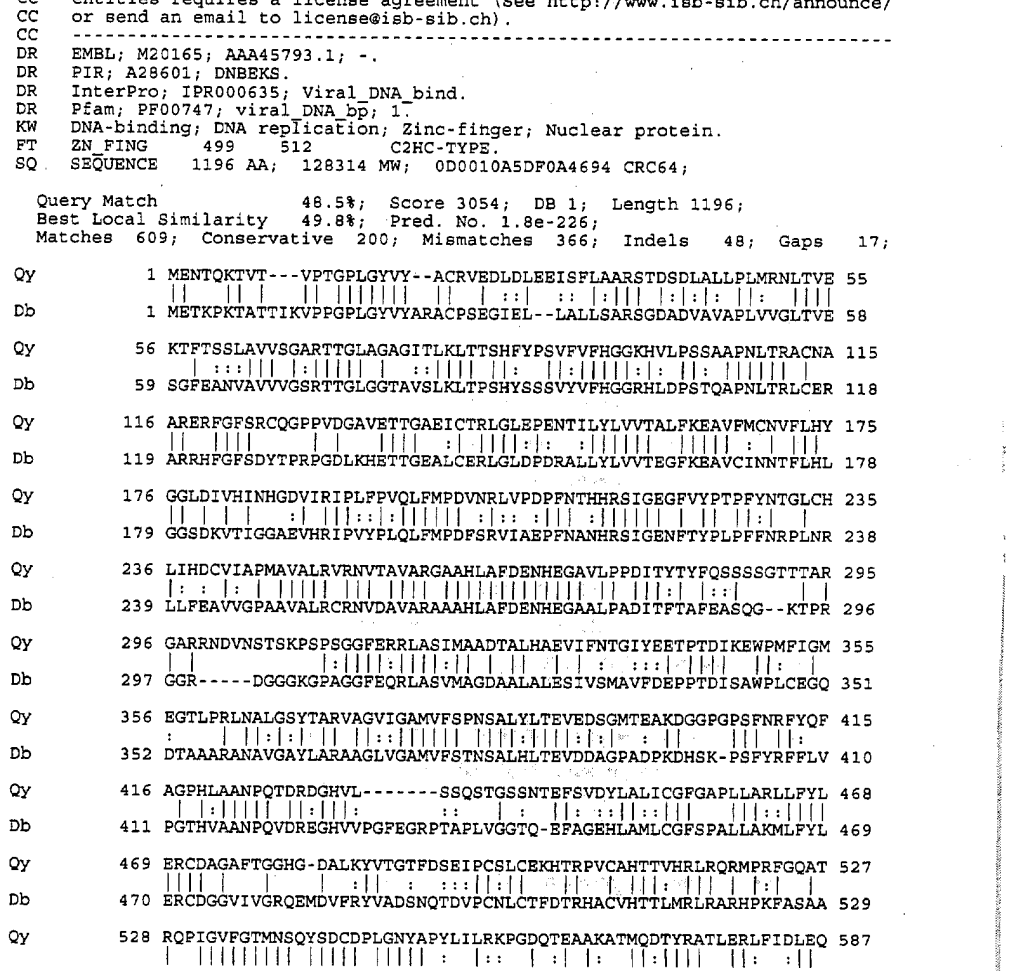
```
RESULT 2
DNBI_HSVEB
ID   DNBI_HSVEB     STANDARD;      PRT;  1209 AA.
AC   P28932;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein.
GN   31.
OS   Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=31520;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92295566; PubMed=1318606;
RA   Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT   "The DNA sequence of equine herpesvirus-1.";
RL   Virology 189:304-316(1992).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86664; AAB02466.1; -.
DR   PIR; E36798; DNBEC4.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING    503    516       C2HC-TYPE.
SQ   SEQUENCE   1209 AA;  129982 MW;  1A728FB04484FE95 CRC64;

  Query Match           57.5%;  Score 3620.5;  DB 1;  Length 1209;
  Best Local Similarity   57.1%;  Pred. No. 4.7e-270;
  Matches  698;  Conservative  191;  Mismatches  299;  Indels   35;  Gaps   14;

Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
              ||:  |||::| |||||||  :    |   :|  |||| |||||:|||:| ||||||||:
Db          1 MESAPKTVSLPVSPLGYVYARQKASLQTGTVSLTAARSVDSDLAVLPVIRGLTVEQTFTT 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
              ::|||:|::|||| | |||||| ||| |: |||:||  :  || |||||||| |||:||
Db         61 NVAVVAGSKTTGLGGTGITLKLTPSHFNPNAFVFYGGSVIGASSNAPNLTRACEAARRRF 120

Qy        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
              |||    |||: |||:|  |||  | | || | |||||   ||| |::|| ||||||
Db        121 GFSAFSSPPVENAVETSGEEICASLNLSPETTALYLVVTESFKEMVYVCNTFLHYGGTST 180

Qy        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
              | |:  | ::|::|||:|||||||  :|||  |||||: ||| ||:|: || |:|
Db        181 VTIDGQDAMKIPIYPVQLYMPDVNRLASEPFNAKHRSIGDEFVYSRPFFNSDLCRLLHGY 240

Qy        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSS--SSGTTTARGAR 298
              |: | |||||||:  |||||||| |||||:||| |:|:|  | |:  ::|  : | |
Db        241 VLGPAAVALRVRNLDGVARGAAHLALDENHEGSVLPQDVTFTLFDSTQGNAGKGSGRAQR 300

Qy        299 RNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGT 358
              : |  | || | | |||||:|||||||  : |  |||:     ::|||:    : |
Db        301 QGD-GSGSKNSASSGIERRLASVMAADTALSVDSIMGAGIYDTELPSVEDWPVLSSGDDT 359

Qy        359 LPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGP 418
              | ||||| ||::|::|||||| || ||:|||:| |  : |||  ||:|||  | |
Db        360 -ESLEALGAYAARLSGLVGAMVFSANSVLYMTEVDDGGPADGKDGS-NPSYHRFYLIAAP 417
```



```
Qy        419 HLAANPQTDRDGHVL------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCD 472
Db        418 YVAGNPQTDKDGRVLPHTADQQAAPINGSNQEFSLDYLALACGFCPQILARLLFYLERCD 477

Qy        473 AGAFTG-GHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPI 531
Db        478 AGTFGGRNETDALRYLANTLESDVPCGLCNQATRPACAHTTLHRLRQRLPRFGAPVRAPI 537

Qy        532 GVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLL 591
Db        538 GIFGTMNSAYSDCDVLGNYASYGALKRPND-NEAPKSIMQDTYRATMERLVNELEQAKLI 596

Qy        592 DR-----GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGL 646
Db        597 DKETLAQASPCSAP--TSVVHDQASFIGLLSNIKDTIEGAAEQFMRTLVEARDFKIREGL 654

Qy        647 SEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQF 706
Db        655 ADANHTMSISLDPYSSSFCPVTSFLARRTVFAVLQDLVLSQCHCLFYGQSVEGRNFRNQF 714

Qy        707 QPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766
Db        715 QPVLRRRFLDMLNGGFITAKTVTVTVSDSGVLAPDLTRPASEPPTKDYDGDMARVSMEVL 774

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826
Db        775 RDLRVKNRVLFSNGGANMSEAARARVAGMASAYRRPDKGSNILNGAVGFLVKQYHGVLFP 834

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886
Db        835 RGHPPGIDTPNPQWFWTLLQRNQMPARLLSKEDIETITAIKRFSDEYSAINFINLTPNNI 894

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
Db        895 GELAQFYFANLVLKYCDHSQYFINGLTAIVVGSRRPRDPAAVLAWIDRTINGAADVEPAA 954

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
Db        955 QEVLQRLGSNPAAWTGTFTSTNMVRYVMDQRPMVVIGLSISKYNGSAGNNRVFQAGNWNG 1014

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
Db       1015 LNGGKNVCPLMAFDRTRRFVLACPRVGFTCEAGGFGTGVRENTLSEQVRGIVSEGGPMVQ 1074

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAV 1126
Db       1075 TAVFAAVLHALGARTQHLAVDDWIGLVDDEFLAASLDALNATVVDQF-GEWSVEAAQELV 1133

Qy       1127 KILDEKTTAG---DGETPTNLAFNFDSC---EPSHDTTSNVLNISGSNISGSTVPGLKRP 1180
Db       1134 KNMEAQTTAGAVAAGEG----AFDFGACVGDTPQQSTSA----FNGGLAMAAAPAGQKRS 1185

Qy       1181 PEDDELFDLSGIPIKHGNITMEM 1203
Db       1186 LPDDILFDMGAPPEKKSGLTFDM 1208
```

```
ESULT 3
NBI_HSV1K
)   DNBI_HSV1K     STANDARD;      PRT;  1196 AA.
;   P17470;
[   01-AUG-1990 (Rel. 15, Created)
[   01-AUG-1990 (Rel. 15, Last sequence update)
?   16-OCT-2001 (Rel. 40, Last annotation update)
E   Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
N   DBP OR UL29 OR ICP8.
3   Herpes simplex virus (type 1 / strain KOS).
:   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
:   Alphaherpesvirinae; Simplexvirus.
(   NCBI_TaxID=10306;
I   [1]
>   SEQUENCE FROM N.A.
```

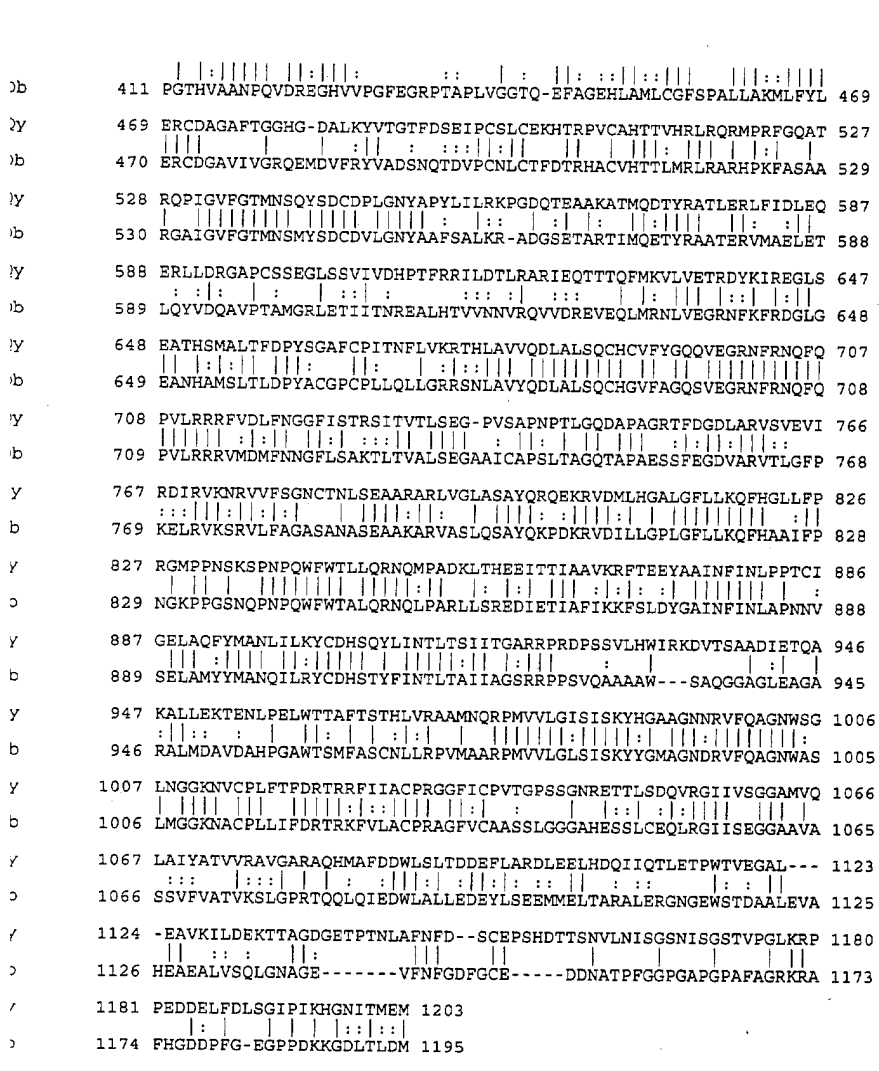
```
RX   MEDLINE=88179536; PubMed=2833010;
RA   Gao M., Bouchey J., Curtin K., Knipe D.M.;
RT   "Genetic identification of a portion of the herpes simplex virus ICP8
RT   protein required for DNA-binding.";
RL   Virology 163:319-329(1988).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC       FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
CC       (PREREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT
CC       MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION
CC       COMPARTMENTS).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20165; AAA45793.1; -.
DR   PIR; A28601; DNBEKS.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     499    512       C2HC-TYPE.
SQ   SEQUENCE   1196 AA;  128314 MW;  0D0010A5DF0A4694 CRC64;

  Query Match           48.5%;  Score 3054;  DB 1;  Length 1196;
  Best Local Similarity  49.8%;  Pred. No. 1.8e-226;
  Matches  609;  Conservative  200;  Mismatches  366;  Indels   48;  Gaps   17;
```



```
Qy          1 MENTQKTVT---VPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVE 55
Db          1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSGDADVAVAPLVVGLTVE 58

Qy         56 KTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNA 115
Db         59 SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCER 118

Qy        116 ARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175
Db        119 ARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINNTFLHL 178

Qy        176 GGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCH 235
Db        179 GGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFNRPLNR 238

Qy        236 LIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTAR 295
Db        239 LLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KTPR 296

Qy        296 GARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355
Db        297 GGR-----DGGGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLCEGQ 351

Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415
Db        352 DTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLV 410

Qy        416 AGPHLAANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYL 468
Db        411 PGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYL 469

Qy        469 ERCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQAT 527
Db        470 ERCDGGVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAA 529

Qy        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
```

```
Db        530 RGAIGVFGTMNSMYSDCDVLGNYAAFSALKR-ADGSETARTIMQETYRAATERVMAELET 588

Qy        588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647

Db        589 LQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLG 648

Qy        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707

Db        649 EANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQ 708

Qy        708 PVLRRRFVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766

Db        709 PVLRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFP 768

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826

Db        769 KELRVKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFP 828

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886

Db        829 NGKPPGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFINLAPNNV 888

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946

Db        889 SELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAW---SAQGGAGLEAGA 945

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006

Db        946 RALMDAVDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWAS 1005

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066

Db       1006 LMGGKNACPLLIFDRTRKFVLACPRAGFVCAASNLGGGAHESSLCEQLRGIISEGGAAVA 1065

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL--- 1123

Db       1066 SSVFVATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVA 1125

Qy       1124 -EAVKILDEKTTAGDGETPTNLAFNFD--SCEPSHDTTSNVLNISGSNISGSTVPGLKRP 1180

Db       1126 HEAEALVSQLGNAGE-------VFNFGDFGCE-----DDNATPFGGPGAPGPAFAGRKRA 1173

Qy       1181 PEDDELFDLSGIPIKHGNITMEM 1203

Db       1174 FHGDDPFG-EGPPDKKGDLTLDM 1195
```

```
RESULT 4
DNBI_HSV11
ID   DNBI_HSV11     STANDARD;      PRT;  1196 AA.
AC   P04296;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
GN   DBP OR UL29 OR ICP8.
OS   Herpes simplex virus (type 1 / strain 17).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88274327; PubMed=2839594;
RA   McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA   McNab D., Perry L.J., Scott J.E., Taylor P.;
RT   "The complete DNA sequence of the long unique region in the genome of
RT   herpes simplex virus type 1.";
RL   J. Gen. Virol. 69:1531-1574(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86067223; PubMed=2999714;
RA   Quinn J.P., McGeoch D.J.;
RT   "DNA sequence of the region in the genome of herpes simplex virus
RT   type 1 containing the genes for DNA polymerase and the major DNA
RT   binding protein.";
RL   Nucleic Acids Res. 13:8143-8163(1985).
RN   [3]
RP   SEQUENCE OF 1062-1196 FROM N.A.
RX   MEDLINE=88306232; PubMed=2457278;
RA   Hammerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
RA   Ludwig H.;
RT   "Common epitopes of glycoprotein B map within the major DNA-binding
RT   proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT   virus type 1 (HSV-1).";
RL   Virology 165:406-418(1988).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC       FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
CC       (PREREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT
CC       MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION
CC       COMPARTMENTS).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X14112; CAA32322.1; -.
DR   EMBL; X03181; CAA26940.1; -.
DR   EMBL; M21631; AAA45787.1; -.
DR   PIR; A03790; DNBEV1.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     499    512       C2HC-TYPE.
SQ   SEQUENCE   1196 AA;  128349 MW;  453799162E5B99E9 CRC64;

  Query Match           48.4%;  Score 3049;  DB 1;  Length 1196;
  Best Local Similarity   49.8%;  Pred. No. 4.4e-226;
  Matches  609;  Conservative  199;  Mismatches  367;  Indels   48;  Gaps   17;

Qy          1 MENTQKTVT---VPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVE 55

Db          1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSGDSDVAVAPLVVGLTVE 58

Qy         56 KTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNA 115

Db         59 SGPEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCER 118

Qy        116 ARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175

Db        119 ARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINNTFLHL 178

Qy        176 GGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCH 235

Db        179 GGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGEKFTYPLPFFNRPLNR 238

Qy        236 LIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTAR 295

Db        239 LLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KTPR 296

Qy        296 GARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355

Db        297 GGR-----DGGGKGAAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLFEGQ 351

Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415

Db        352 DTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLV 410

Qy        416 AGPHLAANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYL 468
```

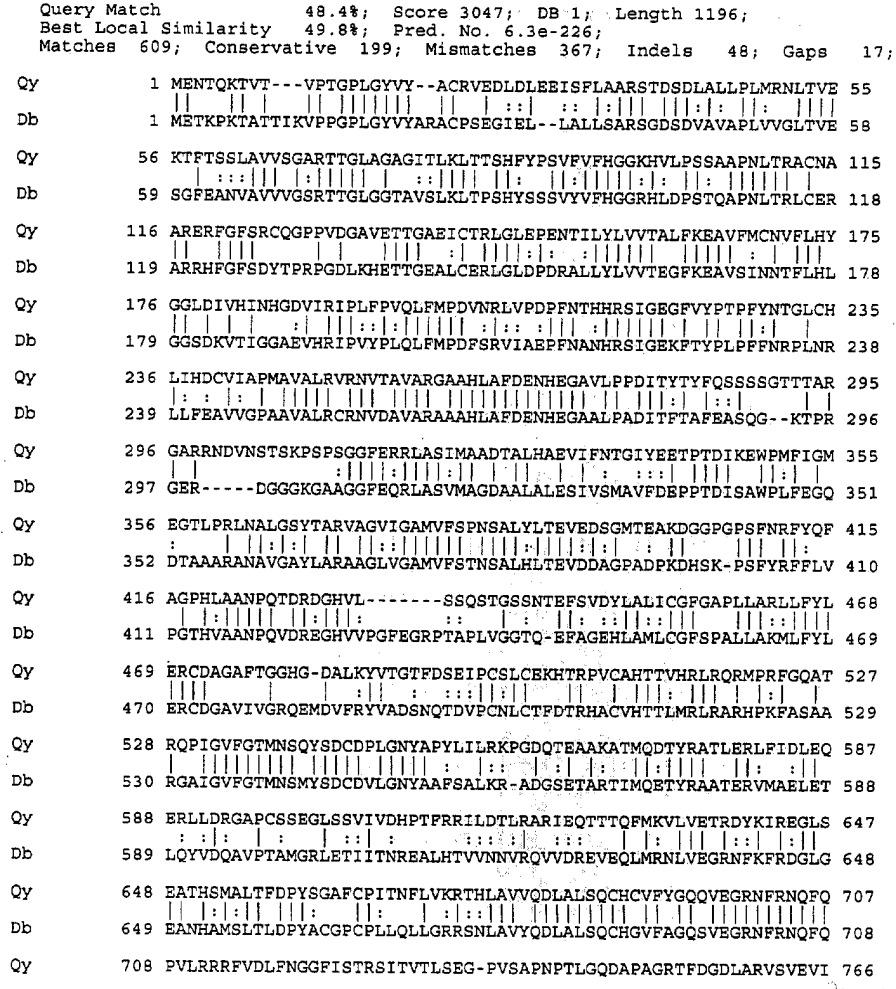


```
)b        411 PGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYL 469
)y        469 ERCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQAT 527
)b        470 ERCDGAVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAA 529
)y        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
)b        530 RGAIGVFGTMNSMYSDCDVLGNYAAFSALKR-ADGSETARTIMQETYRAATERVMAELET 588
)y        588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647
)b        589 LQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLG 648
)y        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707
)b        649 EANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQ 708
)y        708 PVLRRRFVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766
)b        709 PVLRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFP 768
y         767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826
b         769 KELRVKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFP 828
y         827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886
b         829 NGKPPGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKPSLDYGAINFINLAPNNV 888
y         887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
b         889 SELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAW---SAQGGAGLEAGA 945
y         947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
b         946 RALMDAVDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWAS 1005
y        1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
b        1006 LMGGKNACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHESSLCEQLRGIISEGGAAVA 1065
y        1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL--- 1123
b        1066 SSVFVATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVA 1125
y        1124 -EAVKILDEKTTAGDGETPTNLAFNFD--SCEPSHDTTSNVLNISGSNISGSTVPGLKRP 1180
b        1126 HEAEALVSQLGNAGE-------VFNFGDFGCE-----DDNATPFGGPGAPGPAFAGRKRA 1173
y        1181 PEDDELFDLSGIPIKHGNITMEM 1203
b        1174 FHGDDPFG-EGPPDKKGDLTLDM 1195
```

```
ESULT 5
NBI_HSV1F
)   DNBI_HSV1F     STANDARD;      PRT;  1196 AA.
2   P17469;
T   01-AUG-1990 (Rel. 15, Created)
T   01-AUG-1990 (Rel. 15, Last sequence update)
T   16-OCT-2001 (Rel. 40, Last annotation update)
E   Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
N   DBP OR UL29 OR ICP8.
S   Herpes simplex virus (type 1 / strain F).
C   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C   Alphaherpesvirinae; Simplexvirus.
(   NCBI_TaxID=10304;
N   [1]
P   SEQUENCE FROM N.A.
X   MEDLINE=88306231; PubMed=2841793;
A   Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,
```

```
RA   Buhk H.J.;
RT   "Conservation of a gene cluster including glycoprotein B in bovine
RT   herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";
RL   Virology 165:388-405(1988).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC       FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
CC       (PREREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT
CC       MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION
CC       COMPARTMENTS).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
DR   PIR; D29242; DNBEHF.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     499    512       C2HC-TYPE.
SQ   SEQUENCE   1196 AA;  128373 MW;  BC872584DDB1C8E2 CRC64;
```

Query Match 48.4%; Score 3047; DB 1; Length 1196;
Best Local Similarity 49.8%; Pred. No. 6.3e-226;
Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;



```
Qy          1 MENTQKTVT---VPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVE 55
Db          1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSGDSDVAVAPLVVGLTVE 58
Qy         56 KTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNA 115
Db         59 SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCER 118
Qy        116 ARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175
Db        119 ARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVSINNTFLHL 178
Qy        176 GGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCH 235
Db        179 GGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGEKFTYPLPFFNRPLNR 238
Qy        236 LIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTAR 295
Db        239 LLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KTPR 296
Qy        296 GARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355
Db        297 GER-----DGGGKGAAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLFEGQ 351
Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415
Db        352 DTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLV 410
Qy        416 AGPHLAANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYL 468
Db        411 PGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYL 469
Qy        469 ERCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQAT 527
Db        470 ERCDGAVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAA 529
Qy        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
Db        530 RGAIGVFGTMNSMYSDCDVLGNYAAFSALKR-ADGSETARTIMQETYRAATERVMAELET 588
Qy        588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647
Db        589 LQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLG 648
Qy        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707
Db        649 EANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQ 708
Qy        708 PVLRRRFVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766
```

```
                 |||||| :|:|| ||:| :::|| |||| : ||: | || ||| :|:||:|||::
Db        709 PVLRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFP 768

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826
              :::|||:||:|:|    | ||||:||:  | |||| :||||:| | |||||||||  :||
Db        769 KELRVKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFP 828

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886
              | || |  |||||||| |||||:||   |: |:| ||| :|:|: :| ||||||| |  :
Db        829 NGKPPGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFINLAPNNV 888

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
              ||| :|||| ||:||||| | |||||:|| |:||| :    |        | :|  |
Db        889 SELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAW---SAQGGAGLEAGA 945

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
              :||::  :  |  ||: | | :|:|  | |||||||:|||||:| |||:|||||||:
Db        946 RALVDAVDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWAS 1005

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
              | |||| ||| |||||:|::||||| ||:|  :    |  |::| :|:|||| ||| |
Db       1006 LMGGKNACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHESSLCEQLRGIISEGGAAVA 1065

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL--- 1123
               :::   |:::| | | :  :|||:| :||:|: :: ||  : ::    |: : ||
Db       1066 SSVFVATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVA 1125

Qy       1124 -EAVKILDEKTTAGDGETPTNLAFNFD--SCEPSHDTTSNVLNISGSNISGSTVPGLKRP 1180
               ||  :: :   ||:          |||     ||       |     |    |  | ||
Db       1126 HEAEALVSQLGNAGE-------VFNFGDFGCE-----DDNATPFGGPGAPGPAFAGRKRA 1173

Qy       1181 PEDDELFDLSGIPIKHGNITMEM 1203
                |: |   | | |  |::|::|
Db       1174 FHGDDPFG-EGPPDKKGDLTLDM 1195
```

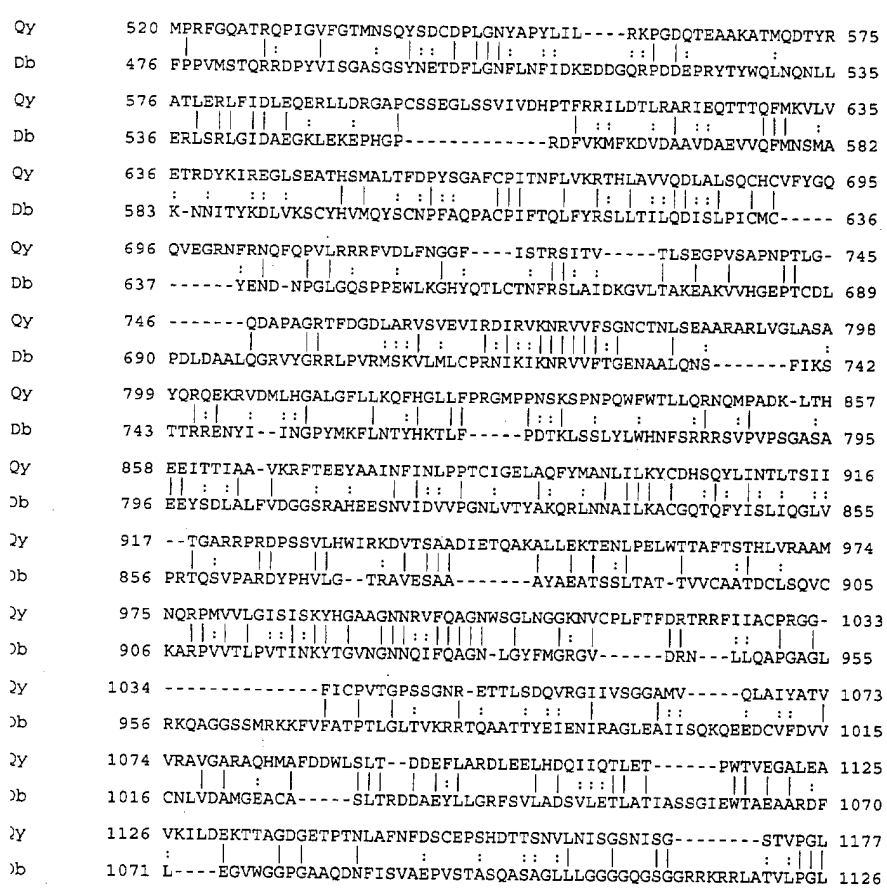
```
RESULT 6
DNBI_HSV2H
ID   DNBI_HSV2H     STANDARD;      PRT;  1196 AA.
AC   P89452;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
GN   DBP OR UL29 OR ICP8.
OS   Herpes simplex virus (type 2 / strain HG52).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10315;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dolan A.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z86099; CAB06754.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     499    512       C2HC-TYPE.
FT   DOMAIN     1168   1196       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
```

```
SQ   SEQUENCE   1196 AA;  128423 MW;  A19CA843280DD7F5 CRC64;

  Query Match           47.9%;  Score 3015.5;  DB 1;  Length 1196;
  Best Local Similarity  49.0%;  Pred. No. 1.7e-223;
  Matches  595;  Conservative  213;  Mismatches  366;  Indels   41;  Gaps   16;

Qy          4 TQKTVTVPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
              | || || ||:|||| || | :|  :| |:||| |:|:: ||:  |||| | ::
Db          7 TTTTVKVPPGPMGYVYGRACPAEGLEL--LSLLSARSGDADVAVAPLIVGLTVESGFEAN 64

Qy         62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFG 121
              :| |:|||| | ::||  ||: |||:|||||:|: ||: |||||| |   || ||
Db         65 VAAVVGSRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQAPNLTRLCERARPHFG 124

Qy        122 FSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIV 181
              |:    |  |||| :| |||||:|: :|||||:  |:||| :  | ||| ||:| |
Db        125 FADYAPRPCDLKHETTGDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGGMDKV 184

Qy        182 HINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCV 241
              |  :| |||::|:|:|||| :|:: |||| :|||||| || ||:|  | |::: |
Db        185 TIGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAV 244

Qy        242 IAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRND 301
              : |||||| ||| |||| |||||||||||||| || ||||:| |::|     ||||
Db        245 VGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KPQRGAR--- 299

Qy        302 VNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPR 361
              :|||||||||||| | || | | : :::| || | | ||  ||  |  | | |
Db        300 --DAGNKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPPDITTWPLLEGQETPAAR 357

Qy        362 LNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLA 421
              |:|:| || |||::|||||| ||||:||||:|:  : ||    || ||:  | |:|
Db        358 AGAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVA 416

Qy        422 ANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAG 474
              |||| ||:|||:       ::  | : ||: ::|||::||  |||::|||||||| |
Db        417 ANPQLDREGHVVPGYEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYLERCDGG 475

Qy        475 AFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGV 533
                    |  :|| :  ::::|:||  ||  |||||: ||| | :|  | |. |||
Db        476 VIVGRQEMDVFRYVADSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGV 535

Qy        534 FGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDR 593
              |||||| |||||:|||||  :  |::  | :|  :  ||:|||  ||  :|| : :|:
Db        536 FGTMNSAYSDCDVLGNYAAFSALKR-ADGSENTRTIMQETYRAATERVMAELEALQYVDQ 594

Qy        594 GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTQFMKVLVETRDYKIREGLSEATHSM 653
              | :  | :::|  :  ::: ::  :::  | | :| :|  | :|:|||  |  |
Db        595 AVPTALGRLETIIGNREALHTVVNNIKQLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654

Qy        654 ALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRR 713
              :|: |||:  ||:  | :|::||| |||||||| || || |||||||||||||||||
Db        655 SLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRRR 714

Qy        714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
              :|||| ||:| :::|| |||| : ||: | || ||| :|:||:|||::  :::|||
Db        715 VMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVK 774

Qy        773 NRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPN 832
              :|:|:|  |||||:|:  | |||| :|||||| | |||| :|||||||||| :|| | |||
Db        775 SRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPPG 834

Qy        833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
              | |||||||| ||||:||  |: :| ||| :|||: :| ||||||| |  : ||| :
Db        835 SNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMY 894

Qy        893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
              ||| ||:|||| |||||:|  |:|||     :  :  |  :|  |:|||:
Db        895 YMANQILRYCDHSTYFINTLTAVIAGSRRPPSVQAAAAWAPQ---GGAGLEAGARALMDS 951

Qy        953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
              : | ||: | :|:|  | |||||:||||:|:|||:|||||||: | ||||
Db        952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGKN 1011
```

```
Qy       1013 VCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYAT 1072
Db       1012 ACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071

Qy       1073 VVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVKI 1128
Db       1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEAL 1131

Qy       1129 LDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFD 1188
Db       1132 VSQLGAAGE-------VFNFGDFGDEDDHAASFGGLAAA--AGAAGVARKRAFHGDDPFG 1182

Qy       1189 LSGIPIKHGNITMEM 1203
Db       1183 -EGPPEKK-DLTLDM 1195


RESULT 7
DNBI_HSV2
ID   DNBI_HSV2      STANDARD;      PRT;  1197 AA.
AC   P36384;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
GN   DBP OR UL29 OR ICP8.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93228441; PubMed=8385914;
RA   Toh Y., Liu Y., Tanaka S., Mori R.;
RT   "Nucleotide sequence of the major DNA-binding protein gene of herpes
RT   simplex virus type 2 and a comparison with the type 1.";
RL   Arch. Virol. 129:183-196(1993).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
DR   PIR; A48350; A48350.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     499    512       C2HC-TYPE.
FT   DOMAIN     1169   1197       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ   SEQUENCE   1197 AA;  128412 MW;  C1576BB5B8865BFB CRC64;

 Query Match             47.9%;  Score 3013;  DB 1;  Length 1197;
 Best Local Similarity   49.0%;  Pred. No. 2.6e-223;
 Matches  595;  Conservative  213;  Mismatches  367;  Indels   40;  Gaps   16;

Qy         4 TQKTVTVPTGPLGYVY--ACRVEDLDLEEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
Db         7 TTTTVKVPPGPMGYVYGRACPAEGLEL--LSLLSARSGDADVAVAPLIVGLTVESGFEAN 64

Qy        62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFG 121
Db        65 VAAVVGSGTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQAPNLTRLCERARRHFG 124

Qy       122 FSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIV 181
Db       125 FSDYAPRPCDLKHETTGDALCERLGLDPDRALLLYLVITEGFREAVCISNTFLHLGGMDKV 184

Qy       182 HINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCV 241
Db       185 TIGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAV 244

Qy       242 IAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRND 301
Db       245 VGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KPQRGAR--- 299

Qy       302 VNSTSKPSPSGGPERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPR 361
Db       300 --DAGNKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPPDITTWPLLEGQETPAAR 357

Qy       362 LNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLA 421
Db       358 AGAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVA 416

Qy       422 ANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAG 474
Db       417 ANPQLDREGHVVPGYEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYLERCDGG 475

Qy       475 AFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGV 533
Db       476 VIVGRQEMDVFRYVADSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGV 535

Qy       534 FGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDR 593
Db       536 FGTMNSAYSDCDVLGNYAAFSALKR-ADGSENTRTIMQETYRAATERVMAELEALQYVDQ 594

Qy       594 GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSM 653
Db       595 AVPTALGRLETIIGNREALHTVVNNIKQLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654

Qy       654 ALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRR 713
Db       655 SLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRRR 714

Qy       714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
Db       715 VMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVK 774

Qy       773 NRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPN 832
Db       775 SRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPPG 834

Qy       833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
Db       835 SNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMY 894

Qy       893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
Db       895 YMANQILRYCDHSTYFINTLTAVIAGSRRPPSVQAAAAWAPQ---GGAGLEAGARALMDS 951

Qy       953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
Db       952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGKN 1011

Qy       1013 VCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYAT 1072
Db       1012 ACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071

Qy       1073 VVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVKI 1128
Db       1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEAL 1131

Qy       1129 LDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFD 1188
Db       1132 VSQLGAAGE-------VFNFGDFGDADDHAASFGGLAAA-AAGAAGVARKRAFHGDDPFG 1183

Qy       1189 LSGIPIKHGNITMEM 1203
Db       1184 -EGPPEKK-DLTLDM 1196


RESULT 8
DNBI_HSVB2
ID   DNBI_HSVB2     STANDARD;      PRT;  1186 AA.
AC   P12639;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
```

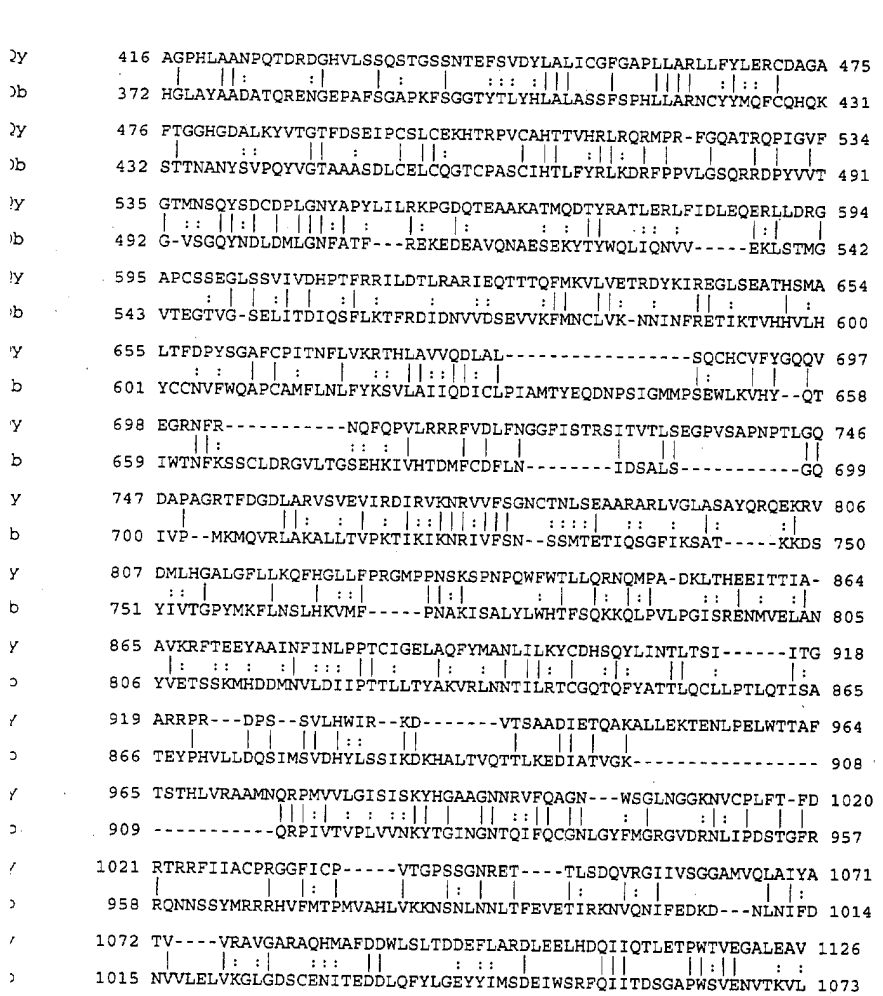
```
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein.
GN   DBP OR UL29.
OS   Bovine herpesvirus type 2 (strain BMV) (Bovine mammillitis virus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88306231; PubMed=2841793;
RA   Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,
RA   Buhk H.-J.;
RT   "Conservation of a gene cluster including glycoprotein B in bovine
RT   herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";
RL   Virology 165:388-405(1988).
RN   [2]
RP   SEQUENCE OF 1058-1186 FROM N.A.
RX   MEDLINE=88306232; PubMed=2457278;
RA   Hammerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
RA   Ludwig H.;
RT   "Common epitopes of glycoprotein B map within the major DNA-binding
RT   proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT   virus type 1 (HSV-1).";
RL   Virology 165:406-418(1988).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21630; AAA46051.1; -.
DR   PIR; A29242; DNBEBG.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     495    508       C2HC-TYPE.
SQ   SEQUENCE   1186 AA;  127286 MW;  A586ECC1479FBD2C CRC64;

  Query Match           46.6%;  Score 2933;  DB 1;  Length 1186;
  Best Local Similarity   48.7%;  Pred. No. 3.7e-217;
  Matches  585;  Conservative  188;  Mismatches  373;  Indels   56;  Gaps   11;

Qy          1 MENTQKTVT---VPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKT 57
              |||  |||  |    |  ||||||||  |:     | |:: |:|||  |||  |:|||:  ||||
Db          1 MENKQKTATTVKVSPGPLGYVYARRLPPEGLTELALLSARSADSDTAVLPLIAGLTVESG 60

Qy         58 FTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAAR 117
              |  ::|||  |:||||: |  |::|||   ||: || :|||||:|: ||||||||:  |: ||
Db         61 FDVNVAVVVGSRTTGVGGTGVSLKLMPSHYAPSAYVFHGGRHLAPSSAAPNLSLLCDRAR 120

Qy        118 ERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGG 177
              :||||  :  | :  ||||  :|  |||  | ::|:|:|:  |||||:: |  || ||
Db        121 VQFGFSSFKPKPCEAEGETTGEALCEHLGLNPNESLLYMVIAEGFKEAVYISNTILHMGG 180

Qy        178 LDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLI 237
              :  |  |   :|:|||::|:|:|||| | |||||  ||:|||| |||| |:|:|  |  |:
Db        181 VGTVTIAGEEVRRIPIYPLQMFMPDYCRAVADPFNDRHRAIGEYFAYPLPFFNAKLASLL 240

Qy        238 HDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGA 297
                 : | |||||  |||  |||||||||||||||||||  |||:|  |  :  |
Db        241 FGAAVGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFDPTGS-----KAG 295

Qy        298 RRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEG 357
               ||          ||||:||||:|| |  ||  | |  :  ::|| ||||  |||    |
Db        296 HRNPREC------GGGFEQRLASVMAGDAALALESIMSMAVFEEPPTDIGTWPMLTCQES 349

Qy        358 TLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAG 417
              |  | ::|:|  | ||:|||||| ||||:|||| :|  : ||   |||  |.  |
Db        350 TAARAASIGAYLGRAAGLVGAMVFSSNSALHLTEVDDAGPADPKD-PTKPSFYRFFLVPG 408

Qy        418 PHLAANPQTDRDGHVLSSQS------TGSSNTEFSVDYLALICGFGAPLLARLLFYLERC 471
              ::|||| |||| |::                | |: :: | :||   |||:::|||||
Db        409 TYVAANPQLDRDGRVVAGHEGRPIVPIVGGNHEFTCEHLATLCGFSPELLAKMLYYLERC 468

Qy        472 DAGAFTG-GHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQP 530
              ||  |  |||:  : :::|| ||   |  |||||: ||| | :|   ||
Db        469 DGGVILGRPEMDTFKYVSDSAHTDVPCCLCSLDNRHSCAHTTLLRLRARHPKFTSTTRGA 528

Qy        531 IGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERL 590
              ||:|| ||| ||||| |||||  :  :::  |  | |:| ||:|||: :||:  :||
Db        529 IGIFGVMNSAYSDCDVLGNYASFSAIKRM-DVQETARAIMQETYRSAVERVMAELENLNY 587

Qy        591 LDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEAT 650
              :|  | |  | |:|      : ::  ::  ::  :  | :||| | :: || | ||
Db        588 IDAAVPTSPAKLESIITGREALQTVVSNVKQVVDGEVAQLMRALVEGRGFRFREALGEAN 647

Qy        651 HSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVL 710
              |:|:|| ||::   ||:   | :|::||| |||||||| || || |||||||:||||
Db        648 HAMSLTLDPHASVPCPLLQMLGRRSNLAVYQDLALSQCHGVFEGQAVEGRNFRSQFQPVL 707

Qy        711 RRRFVDLFNGGFISTRSITVTLSEGP-VSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDI 769
              ||| :|:|| ||:| |::|  |::|  :|||   || | |   :|||:|||::  :: |
Db        708 RRRVLDMFNNGFLSARTLTVALTDGACISAPGLVSGQHAAAESGFEGDVARVNLGFPKEI 767

Qy        770 RVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGM 829
              |||:|||:|  |||||| :|||:|| |||| :|||:|| | |||:|||||  ||| |
Db        768 RVKSRVLFAGAGPAASEAARARIAGLQSAYQKSDKRVDILLGPLGFMLKQFHATLFPNGK 827

Qy        830 PPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGEL 889
              || | :|||||||| ||||:||  | :|:|: || |||| || | ||:|||  | ||
Db        828 PPGSDNPNPQWFWTALQRNQLPARLLSREDISLIAFVKRFSVEYGAGNFVNLPPNNISEL 887

Qy        890 AQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKAL 949
              | :|||| |||||||| |||||:|| :|||  :  |  :  |  ::| ||:::
Db        888 AMYYMANQILKYCDHSTYFINTLTALIAGSRRPPNAQAAAAWAPRGGT---ELEAQARSV 944

Qy        950 LEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNG 1009
              :  : | ||| | | :|:|  | |||||||:|||||:| |||:|||||| : | ||
Db        945 VANPGDHPGAWTTMFASCNLLRPVMATRPMVVLGLSISKYYGMAGNDRVFQAGNLANLLG 1004

Qy       1010 GKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAI 1069
              ||| ||| |||||:|:||||| |||:|        ||  |::| :|:| || ||| |  :
Db       1005 GKNACPLLIFDRTRKFVIACPRAGFVCAAVSAGSGAHESSLCEQLRAIIAEGGATVASDV 1064

Qy       1070 YATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKIL 1129
              :|  ::::||| |  :|||:|:|:|: :: || |  :  ::        |::: ||:: :
Db       1065 FAAAAKSLGARVQQLQIEDWLALLEDEYLSEEMMELAGRALERGGGEWSLDAALDVAREA 1124

Qy       1130 DEKTT--------------AGDGETPTNLAFNFDS--------------CEPSHDTTSN 1160
              :  |                 ||     ||  : |                 | :| |:
Db       1125 EAMVTRHVDAEETFDFGAFAEDGPADAGLAVHLQSRRRPLACSDLFGDAPAEKRNDLTLD 1184

Qy       1161 VL 1162
              :|
Db       1185 ML 1186


RESULT 9
DNBI_HSVE1
ID   DNBI_HSVE1     STANDARD;      PRT;   375 AA.
AC   Q03444;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein (Fragment).
GN   31.
OS   Equine herpesvirus type 1 (isolate HVS25A) (EHV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
```

```
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10327;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94106109; PubMed=8279122;
RA   Bell C.W., Whalley J.M.;
RT   "Herpesvirus ICP18.5 and DNA-binding protein genes are conserved in
RT   equine herpesvirus-1.";
RL   Virus Genes 7:219-228(1993).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13930; BAA03033.1; -.
DR   PIR; JQ0846; JQ0846.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Nuclear protein.
FT   NON_TER       1       1
SQ   SEQUENCE   375 AA;  40309 MW;  ECF327925EBF999B CRC64;

 Query Match             17.4%;  Score 1094.5;  DB 1;  Length 375;
 Best Local Similarity   55.5%;  Pred. No. 1.9e-76;
 Matches  212;  Conservative   57;  Mismatches   98;  Indels   15;  Gaps    5;

Qy       828 GMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIG 887
            |  ||    :|||||||||||||||||  |: |:| || |:|||::||:|||||||  |  ||
Db         2 GHPPGIDTPNPQWFWTLLQRNQMPARLLSKEDIETITAIKRFSDEYSAINFINLTPNNIG 61

Qy       888 ELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAK 947
            |||||| ||||:|||||||||| || ||:|: |:||||||::|| || : :  |||:|  |:
Db        62 ELAQFYFANLVLKYCDHSQYFINGLTAIVVGSRRPRDPAAVLAWIDRTINGAADVEPAAQ 121

Qy       948 ALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGL 1007
             :|::  : |  ||  ||||::||  |:||||||:|:|||||:|:|||||||||||||:||
Db       122 EVLQRLGSNPAAWTGTFTSTNMVRYVMDQRPMVVIGLSISKYNGSAGNNRVFQAGNWNGL 181

Qy      1008 NGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQL 1067
            |||||||||  |||||||::|||| || |   |  :| || |||:||||||:  || |||
Db       182 NGGKNVCPLMAFDRTRRFVLACPRVGFTCEAGGFGTGVRENTLSEQVRGIVSEGGPMVQT 241

Qy      1068 AIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVK 1127
             |::| |: |:||| ||:| |||: | ||||||  |: |:  ::       |:|| | | ||
Db       242 AVFAAVLHALGARTQHLAVDDWIGLVDDEFLAASLDALNATVVDQF-GEWSVEAAQELVK 300

Qy      1128 ILDEKTTAG---DGETPTNLAFNFDSC---EPSHDTTSNVLNISGSNISGSTVPGLKRPP 1181
             :: :||||     ||      ||:| :|    |   |::     :|     :  | ||
Db       301 NMEAQTTAGAVAAGEG----AFDFGACVGDTPQQSTSA----FNGGLAMAAAPAGQKRSL 352

Qy      1182 EDDELFDLSGIPIKHGNITMEM 1203
             || |||:   | |   :| :|
Db       353 PDDILFDMGAPPEKKSGLTFDM 374


RESULT 10
VBI_EBV
ID   DNBI_EBV       STANDARD;      PRT;  1128 AA.
AC   P03227;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein.
GN   BALF2.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01555; CAA24808.1; -.
DR   PIR; A43045; QQBE47.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein;
KW   Early protein.
SQ   SEQUENCE   1128 AA;  123121 MW;  283D5258031A16D6 CRC64;

  Query Match             10.6%;  Score 669;  DB 1;  Length 1128;
  Best Local Similarity   23.2%;  Pred. No. 5.8e-43;
  Matches  292;  Conservative  186;  Mismatches  537;  Indels  246;  Gaps   45;

Qy        13 GPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGARTTG 72
             || ||:|    :    | |:: |              :||:  :|||  |: :   |
Db        18 GPCGYIYFYPLATYPLREVATLGTGYAGHRCLTVPLLCGITVEPGFSIN---VKALHRRP 74

Qy        73 LAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFS------RCQ 126
                |:   |  : ::  ::|||    | | | |  |   || ||:       |
Db        75 DPNCGL---LRATSYHRDIYVFHNAHMVPPIFEGPGLEALCGETREVFGYDAYSALPRES 131

Qy       127 GPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINHG 186
             | |   |              ||:|  :  : :|  |||::  :|       | :
Db       132 SKPGDFFPE----------GLDPSAYLGAVAITEAFKERLYSGNLVAIPSLKQEVAVGQS 181

Qy       187 DVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMA 246
               :|:||:  ::|   | :|                  |||: |  :|: :  :|
Db       182 ASVRVPLYDKEVFPEGVPQL------------------RQFYNSDLSRCMHEALYTGLA 222

Qy       247 VALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTS 306
             |||||  |  :                  :    ||       |: |   :  :::
Db       223 QALRVRRVGKL--------------------VELLEKQSLQDQAKVAKVAPLKEFPAST 261

Qy       307 KPSPSGGFERRLASIMAADT-ALHAEVIFNTGIYE---ETPTDIK--EWPMFIGMEGTLP 360
             |  |      ::|  |: |    | :  :  |  |||  :    ||:|   ||
Db       262 ISHPDSG------ALMIVDSAACELAVSYAPAMLEASHETPASLNYDSWPLFADCEGPEA 315

Qy       361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAG--- 417
             |: ||  | | :|  :  :|:|: || ||:     ||::: |  || ||  |
Db       316 RVAALHRYNASLAPHVSTQIFATNSVLYV-----SGVSKSTGQGKESLFNSFYMTHGLGT 370

Qy       418 -----------PHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLF 466
                      |  :   |  |     :  |  |  ::|::|    |  ||||  :
Db       371 LQEGTWDPCRRPCFSGWGGPDVTG------TNGPGN--YAVEHLVYAASFSPNLLARYAY 422

Qy       467 YLERCDAGAFTGGHGDAL-------KYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQR 519
             ||: |      | :|          || |   | : |||||     || :|   ||| |
Db       423 YLQFCQ------GQKSSLTPVPETGSYVAGAAASPM-CSLCEGRAPAVCLNTLFFRLRDR 475
```



```
Qy        520 MPRFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLIL----RKPGDQTEAAKATMQDTYR 575
Db        476 FPPVMSTQRRDPYVISGASGSYNETDFLGNFLNFIDKEDDGQRPDDEPRYTYWQLNQNLL 535

Qy        576 ATLERLFIDLEQERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLV 635
Db        536 ERLSRLGIDAEGKLEKEPHGP------------RDFVKMFKDVDAAVDAEVVQFMNSMA 582

Qy        636 ETRDYKIREGLSEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQ 695
Db        583 K-NNITYKDLVKSCYHVMQYSCNPFAQPACPIFTQLFYRSLLTILQDISLPICMC----- 636

Qy        696 QVEGRNFRNQFQPVLRRRFVDLFNGGF----ISTRSITV-----TLSEGPVSAPNPTLG- 745
Db        637 ------YEND-NPGLGQSPPEWLKGHYQTLCTNFRSLAIDKGVLTAKEAKVVHGEPTCDL 689

Qy        746 -------QDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASA 798
Db        690 PDLDAALQGRVYGRRLPVRMSKVLMLCPRNIKIKNRVVFTGENAALQNS-------FIKS 742

Qy        799 YQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQRNQMPADK-LTH 857
Db        743 TTRRENYI--INGPYMKFLNTYHKTLF-----PDTKLSSLYLWHNFSRRRSVPVPSGASA 795

Qy        858 EEITTIAA-VKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSII 916
Db        796 EEYSDLALFVDGGSRAHEESNVIDVVPGNLVTYAKQRLNNAILKACGQTQFYISLIQGLV 855

Qy        917 --TGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAM 974
Db        856 PRTQSVPARDYPHVLG--TRAVESAA-------AYAEATSSLTAT-TVVCAATDCLSQVC 905

Qy        975 NQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGG- 1033
Db        906 KARPVVTLPVTINKYTGVNGNNQIFQAGN-LGYFMGRGV------DRN---LLQAPGAGL 955

Qy       1034 --------------FICPVTGPSSGNR-ETTLSDQVRGIIVSGGAMV-----QLAIYATV 1073
Db        956 RKQAGGSSMRKKFVFATPTLGLTVKRRTQAATTYEIENIRAGLEAIISQKQEEDCVFDVV 1015

Qy       1074 VRAVGARAQHMAFDDWLSLT--DDEFLARDLEELHDQIIQTLET------PWTVEGALEA 1125
Db       1016 CNLVDAMGEACA-----SLTRDDAEYLLGRFSVLADSVLETLATIASSGIEWTAEAARDF 1070

Qy       1126 VKILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISG--------STVPGL 1177
Db       1071 L----EGVWGGPGAAQDNFISVAEPVSTASQASAGLLLGGGGQGSGGRRKRRLATVLPGL 1126

Qy       1178 K 1178
Db       1127 E 1127
```

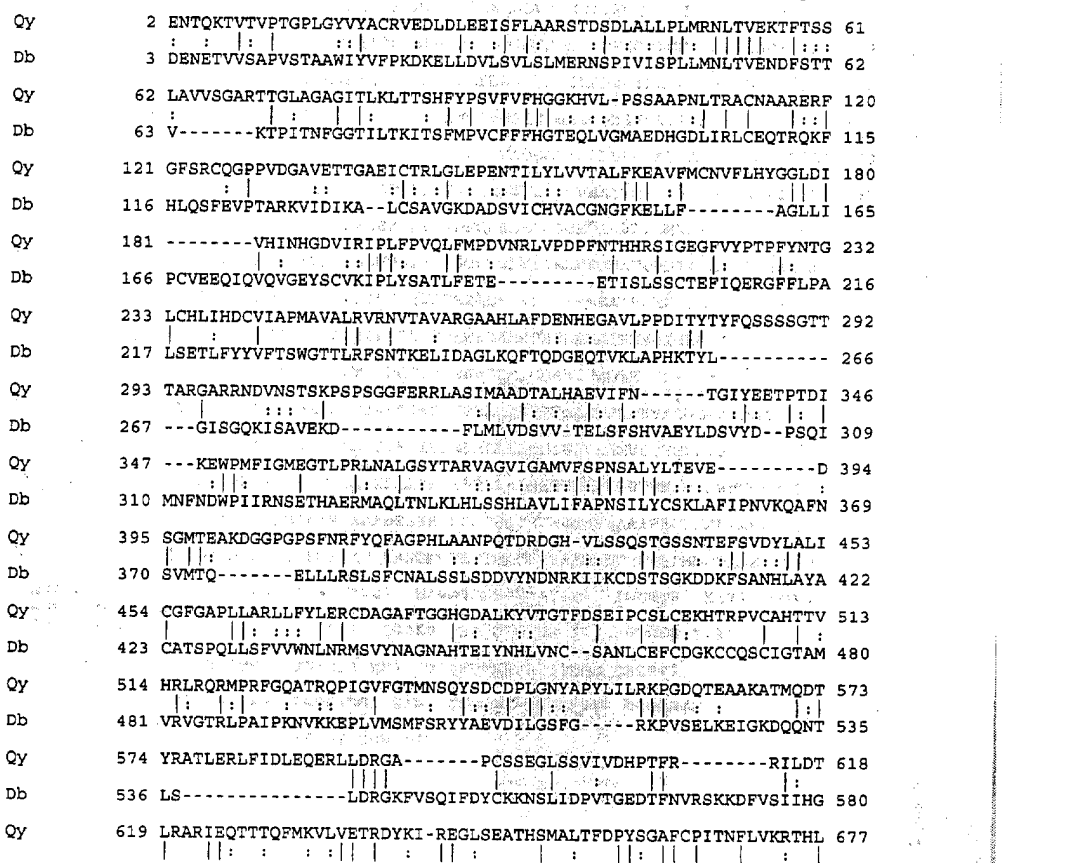
```
RESULT 11
NBI_HSVSA
ID   DNBI_HSVSA     STANDARD;      PRT;  1128 AA.
AC   P24910;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein.
GN   6 OR KFRF1.
OS   Herpesvirus saimiri (strain 11).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=10383;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90163221; PubMed=2154888;
RA   Albrecht J.-C., Fleckenstein B.;
RT   "Structural organization of the conserved gene block of Herpesvirus
RT   saimiri coding for DNA polymerase, glycoprotein B, and major DNA
RT   binding protein.";
RL   Virology 174:533-542(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92333688; PubMed=1321287;
RA   Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA   Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA   Honess R.W.;
RT   "Primary structure of the herpesvirus saimiri genome.";
RL   J. Virol. 66:5047-5058(1992).
RN   [3]
RP   SEQUENCE OF 1-81 FROM N.A.
RX   MEDLINE=91202588; PubMed=1850023;
RA   Nicholas J., Coles L.S., Newman C., Honess R.W.;
RT   "Regulation of the herpesvirus saimiri (HVS) delayed-early 110-
RT   kilodalton promoter by HVS immediate-early gene products and a
RT   homolog of the Epstein-Barr virus R trans activator.";
RL   J. Virol. 65:2457-2466(1991).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64346; CAA45629.1; -.
DR   EMBL; M31122; AAA46162.1; -.
DR   EMBL; M60849; AAA46157.1; -.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     453    466       C2HC-TYPE.
SQ   SEQUENCE   1128 AA;  127457 MW;  BA3ADE861381328D CRC64;

  Query Match             9.6%;  Score 604;  DB 1;  Length 1128;
  Best Local Similarity   22.9%;  Pred. No. 5.9e-38;
  Matches  275;  Conservative  189;  Mismatches  509;  Indels  226;  Gaps   40;

Qy         11 PTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGART 70
Db         18 PVEPCGYIYVYPKEGFPFKEASLLGNKNVGASAMSLPLLSDLTVESNFSFNVKAVHKKID 77

Qy         71 TGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPPV 130
Db         78 M------TTLLVRVSAYHREAIVFFNTDLFEPIFVGPGLDILCSDARSLFGYTNF--VPR 129

Qy        131 DGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCN-VFLHYGGLDIVHINHGDVI 189
Db        130 TDLRDTVDIKDLYAPFYSEDSCFMAVVVTEGFKERLYFGNLVPIIAQGLK-VQINGREAV 188

Qy        190 RIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVAL 249
Db        189 KIPLYDEDLFSKSHEHL------------------PRFYIPSVSKYLHDSVFTSIAQAL 229

Qy        250 RVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPS 309
Db        230 RIRDVESVIRASEKQSIQDQYKLAKI-------------------------VNSK---- 259

Qy        310 PSGGFERRLASIMAADTA-------LHAEVIFNTGI-YEETP------TDIKEWPMFIGM 355
Db        260 -----DFSLQSVKCQDASAFMVIDCIAAELAMSYGLSFLEAPQDPCAVLDYTSWPIFETA 314

Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415
Db        315 ETEEDRIKAIQDWNAMMSVHVYTHLFSTNSVLYLTKINKQTQSNKSEQN---VYNTYFMQ 371
```

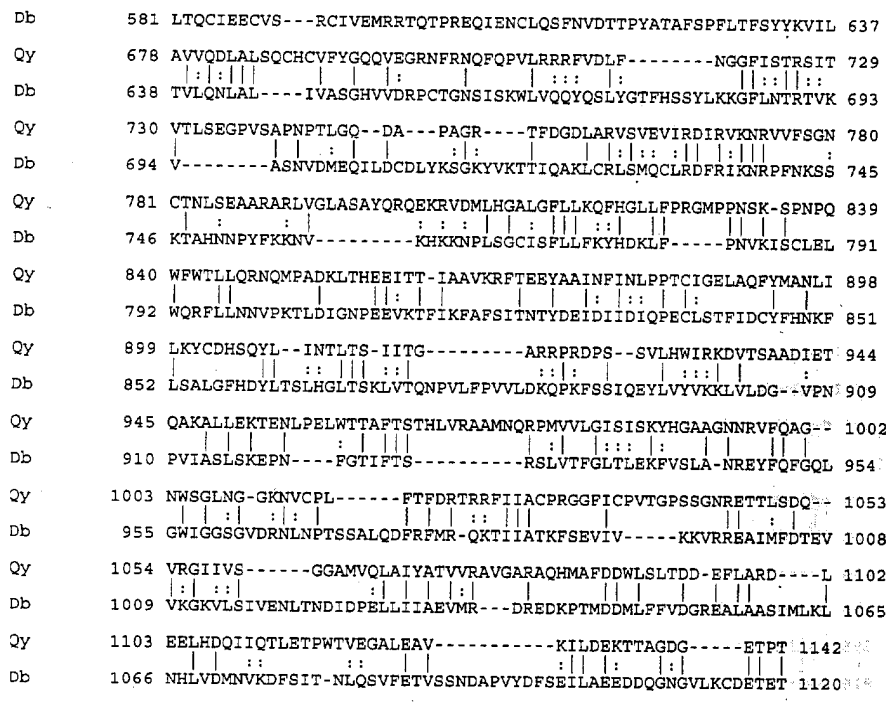


ESULT 12
NBI_HSV6U
D DNBI_HSV6U STANDARD; PRT; 1132 AA.
C P52338;
D 01-OCT-1996 (Rel. 34, Created)
D 01-OCT-1996 (Rel. 34, Last sequence update)
D 16-OCT-2001 (Rel. 40, Last annotation update)
E Major DNA-binding protein (MDBP).
N U41.
S Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
C Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Betaherpesvirinae; Roseolovirus.
C NCBI_TaxID=10370;
R [1]
P SEQUENCE FROM N.A.
X MEDLINE=95266321; PubMed=7747482;
A Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
A Martin M.E., Efstathiou S., Craxton M., Macaulay H.A.;
T "The DNA sequence of human herpesvirus-6: structure, coding content,
T and genome evolution.";

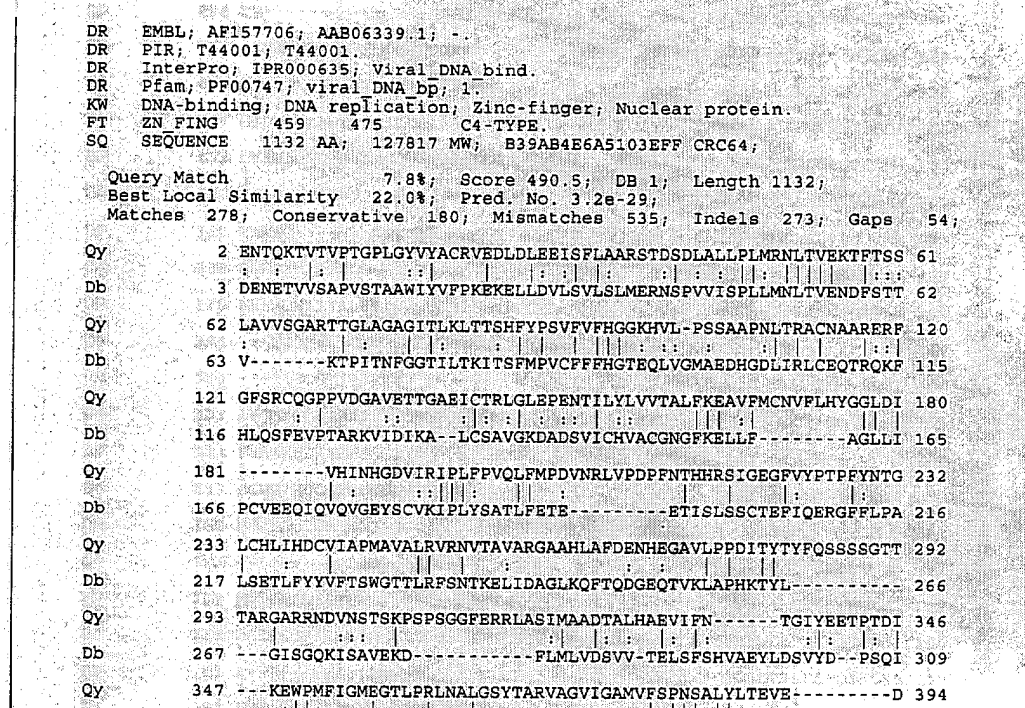
RL Virology 209:29-51(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X92436; CAA63167.1; -.
DR EMBL; X83413; CAA58375.1; -.
DR InterPro; IPR000635; Viral_DNA_bind.
DR Pfam; PF00747; viral_DNA_bp; 1.
KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN_FING 459 475 C4-TYPE.
SQ SEQUENCE 1132 AA; 127762 MW; 4468D3E5559F02D8 CRC64;

Query Match 7.8%; Score 492.5; DB 1; Length 1132;
Best Local Similarity 21.8%; Pred. No. 2.3e-29;
Matches 274; Conservative 185; Mismatches 544; Indels 253; Gaps 53;





```
RESULT 13
DNBI_HSV6Z
ID   DNBI_HSV6Z     STANDARD;      PRT;  1132 AA.
AC   P52538;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein (MDBP).
GN   U41 OR KA2L.
OS   Human herpesvirus (type 6 / strain Z29) (HHV6).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Roseolovirus.
OX   NCBI_TaxID=36351;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95074921; PubMed=7983761;
RA   Stamey F.R., Dominguez G., Black J.B., Dambaugh T.R., Pellett P.E.;
RT   "Intragenomic linear amplification of human herpesvirus 6B oriLyt
RT   suggests acquisition of oriLyt by transposition.";
RL   J. Virol. 69:589-596(1995).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```



```
DR   EMBL; AF157706; AAB06339.1; -.
DR   PIR; T44001; T44001.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT   ZN_FING     459    475       C4-TYPE.
SQ   SEQUENCE   1132 AA;  127817 MW;  B39AB4E6A5103EFF CRC64;

  Query Match           7.8%;  Score 490.5;  DB 1;  Length 1132;
  Best Local Similarity   22.0%;  Pred. No. 3.2e-29;
  Matches  278;  Conservative  180;  Mismatches  535;  Indels  273;  Gaps   54;
```

```
Db        782 PNVKISCLELWQRFLLNNVPKTLDIGNPEEVKTFIKFAFSITNTYDEIDIIDIQPECLST 841

Qy        889 LAQFYMANLILKYCDHSQYL--INTLTS-IITG---------ARRPRDPS--SVLHWIRK 934

Db        842 FIDCYFHNKFLSALGFHDYLTSLHGLTSKLVTQNPVLFPVVLDKQPKFSSIQEYLVYVKK 901

Qy        935 DVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAG 994

Db        902 LVLDG--VPNPVIASLSKEPN----FGTIFTS----------RSLVTFGLTLEKFVSLA- 944

Qy        995 NNRVFQAG--NWSGLNG-GKNVCPL------FTFDRTRRFIIACPRGGFICPVTGPSSGN 1045

Db        945 NREYFQFGQLGWIGGSGVDRNLNPTSSALQDFRFMR-QKTIIATKFSEVIV-----KKVR 998

Qy       1046 RETTLSDQ--VRGIIVS------GGAMVQLAIYATVVRAVGARAQHMAFDDWLSLTDD-E 1096

Db        999 RETIMFDTEVVKGKVLSIVENLTNDIDPELLIIAEVMR---DREDKPTMDDMLFFVDGRE 1055

Qy       1097 FLARD----LEELHDQIIQTLETPWTVEGALEAV----------KILDEKTTAGDG--- 1138

Db       1056 ALAASIMLKLNHLVDMNVRDFSIA-NLQSVFEAVSSNDAPVYDFSEILAEEDDQASGVLK 1114

Qy       1139 --ETPT 1142

Db       1115 CDETET 1120
```

```
RESULT 14
NBI_HSV7J
D   DNBI_HSV7J     STANDARD;      PRT;  1131 AA.
C   P52339;
T   01-OCT-1996 (Rel. 34, Created)
T   01-OCT-1996 (Rel. 34, Last sequence update)
T   16-OCT-2001 (Rel. 40, Last annotation update)
E   Major DNA-binding protein (MDBP).
N   U41.
S   Human herpesvirus (type 7 / strain JI) (HHV7).
C   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C   Alphaherpesvirinae; Simplexvirus.
X   NCBI_TaxID=57278;
N   [1]
P   SEQUENCE FROM N.A.
C   STRAIN=JI;
A   Nicholas J.;
L   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
C   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
C       REPLICATION.
C   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
C   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
C       FAMILY.
C   ---------------------------------------------------------------------
C   This SWISS-PROT entry is copyright. It is produced through a collaboration
C   between the Swiss Institute of Bioinformatics and the EMBL outstation -
C   the European Bioinformatics Institute. There are no restrictions on its
C   use by non-profit institutions as long as its content is in no way
C   modified and this statement is not removed. Usage by and for commercial
C   entities requires a license agreement (See http://www.isb-sib.ch/announce/
C   or send an email to license@isb-sib.ch).
C   ---------------------------------------------------------------------
R   EMBL; U43400; AAC54703.1; -.
R   PIR; T41943; T41943.
R   InterPro; IPR000635; Viral_DNA_bind.
R   Pfam; PF00747; viral_DNA_bp; 1.
W   DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
T   ZN_FING    459    475       C2HC-TYPE.
Q   SEQUENCE   1131 AA;  129008 MW;  52C97388D1B6D04F CRC64;

 Query Match             7.7%;  Score 483;  DB 1;  Length 1131;
 Best Local Similarity   20.8%;  Pred. No. 1.2e-28;
 Matches  258;  Conservative  204;  Mismatches  560;  Indels  220;  Gaps   46;

Y         2 ENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
```

```
Db          3 DDNETVVSAPICTAAWLYILPKEQKLIEILTTLSLMEKRKSVVISPLLLNLTVENDFFPT 62

Qy         62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSA-APNLTRACNAARERF 120

Db         63 V-------KTPIINYGGTVITKITSFMPVCFFFHGTDVFLKEAEDHGNLDKLCKQTREKF 115

Qy        121 GFSRCQGPPVDGAVETTG-AEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLD 179

Db        116 NL---QEFVVNGNRKSVDIGKICESVGRNADDVLCHIVVGNGFKELLF--------AGLL 164

Qy        180 I------VHINHGD--VIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNT 231

Db        165 IPCVEEQIQVQVGECLAIKIPLYSATLFESEETLCI---------DTCTEFIQENGFYAP 215

Qy        232 GLCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGT 291

Db        216 QISEVLFYLIFTSWGMTLRFNNTLELIKAGLKQFIQDTEQTVKLAPNKTY---------- 265

Qy        292 TTARGARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFN------TGIYEETP-T 344

Db        266 ---HGIPGQKLSPIEKD----------HLMLVD-AVITELTFSYTAEYLDSIYENNQIM 310

Qy        345 DIKEWPMFIGMEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGG 404

Db        311 NFSEWPIIKSAETHEEKIVELKKLRLHLSSHVAALVFAANSILYSNKLAYISNTK----- 365

Qy        405 PGPSFN---------RFYQFAGPHLAANPQTDRDGH-VLSSQSTGSSNTEFSVDYLALIC 454

Db        366 --QAFNSAITQETLLRSIQFCNSLSSLNEDFYNDARKLIKCNSSPCKEDKFSAFHLAYAC 423

Qy        455 GFGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVH 514

Db        424 ATCPQILSHIIWNLNRMSIYNTNCGNSEIYNHIVNC--SSNLCEFCEGKCCHSCIGTALI 481

Qy        515 RLRQRMPRFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTY 574

Db        482 RINSRLPQISKTTKKEPIVMTMFSRFYADVDVLGSFGKKGVNESKDPMKEAQTTPSLDRF 541

Qy        575 RATLERLFIDLEQERLLDRGAPCSSEGLSSVIV--DHPTFR------RILDTLRARIEQT 626

Db        542 K------FLGMIHDY-------CKKNNLIDAITGEDNLNFKSQNDFVNMINDLIQCIEEA 588

Qy        627 TTQFMKVLVETRDYKIREGLSEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALS 686

Db        589 VSKCISEMRKTQ--TSREQIENCLQSFNIDTTPLSLAFSPFFVFTYYKVILIVLQNLAL- 645

Qy        687 QCHCVFYGQQVE--------GRNFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS 738

Db        646 ---IIGTGYVVDRPCTGNLISKWLMQQYQSLYGAFYNSHFKKGFLNMKTVKI-------- 694

Qy        739 APNPTLGQ---------DAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAAR 789

Db        695 ASNVDMEQYIDFNLFKSGKYAKTSIQAKLCRLSMQCLKDFRVKNRPFNKPNKNTQNN--- 751

Qy        790 ARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQRNQ 849

Db        752 -------PFFKKVKQKKNPLSGCLSFLLFKYHERLF-----PNLKISCLE-FWQRILLNN 798

Qy        850 MP--ADKLTHEEITT-IAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQ 906

Db        799 MPKTIDIGNVEDMRSPIKFTFRVTNSYDEIDLLDIQPECLLSFIEYYFHNKLLSVLGYRD 858

Qy        907 YL--INTLTSIITGARRPRDPSSVLHWIRKDVT--SAADIETQAKALLEKTENLPELWTT 962

Db        859 YLTSLHALTSKLV----PQNPMLFPVFLKEHPTFSSVQEYVMHVKKLV--GNGLKEPMTA 912

Qy        963 AFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNN--RVFQAGNWSGLNGGKNVCPL---- 1016

Db        913 SLTKEPNFGSIFTGRSIITFGLMIEKFVSVASRDYFHFGQLGWIAGSGVDRNLNPPSSGL 972

Qy       1017 --FTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSD--QVRGII------VSGGAMVQ 1066
```

```
Db        973 QDFRFMR-QKFVIATK----LCDII-VKKVKREAIVYDVEVIRGKVLNIIESLSNSVNPE 1026

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDD-EFLARDLE---------ELHDQIIQTLETP 1116

Db       1027 LLILAEVMK---DRDSKPTMDDMLFYVDGREPLAKSVMNKIQHLTDLNVHDFSLSTLLSV 1083

Qy       1117 W---------------TVEGALEAVKILDEKTTAGDGETPT 1142

Db       1084 FEEQVEDSAAIYDFSELLVEGNEQGFGILKCEETEHENEEPS 1125


RESULT 15
DNBI_SCMVC
ID   DNBI_SCMVC     STANDARD;      PRT;  1160 AA.
AC   P13215;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Major DNA-binding protein (MDBP).
GN   UL57 OR DBP.
OS   Simian cytomegalovirus (strain Colburn).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=50292;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91037979; PubMed=2172458;
RA   Anders D.G.;
RT   "Nucleotide sequence of a cytomegalovirus single-stranded DNA-binding
RT   protein gene: comparison with alpha- and gammaherpesvirus
RT   counterparts reveals conserved segments.";
RL   J. Gen. Virol. 71:2451-2456(1990).
RN   [2]
RP   SEQUENCE OF 205-308 FROM N.A.
RX   MEDLINE=88155776; PubMed=2831398;
RA   Anders D.G., Gibson W.;
RT   "Location, transcript analysis, and partial nucleotide sequence of
RT   the cytomegalovirus gene encoding an early DNA-binding protein with
RT   similarities to ICP8 of herpes simplex virus type 1.";
RL   J. Virol. 62:1364-1372(1988).
CC   -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC       REPLICATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00750; BAA00647.1; -.
DR   EMBL; M19868; AAA46066.1; ALT_SEQ.
DR   InterPro; IPR000635; Viral_DNA_bind.
DR   Pfam; PF00747; viral_DNA_bp; 1.
KW   DNA-binding; DNA replication; Zinc-finger; Nuclear protein;
KW   Early protein.
FT   ZN_FING     465    479       C2HC-TYPE.
SQ   SEQUENCE   1160 AA;  129005 MW;  7037716816974B1A CRC64;

  Query Match             7.6%;  Score 479.5;  DB 1;  Length 1160;
  Best Local Similarity   21.3%;  Pred. No. 2.3e-28;
  Matches  279;  Conservative  202;  Mismatches  571;  Indels  255;  Gaps   53;

Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60

Db          1 MSNEELSALAPVGPAAYVYFTKTNHEMNEVLATLSLCDSSSPVVIAPLLMGLTVDQDFCT 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLP-SSAAPNLTRACNAARER 119

Db         61 SV------RTPVVCYDGGVLTKVTS-FCPFALYFYNTQGIVDFSEPHGDVQRLCDETRQR 113

Qy        120 FGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLD 179

Db        114 YAIESYM--PEEGRAPTDLAALCTAAGCDPQEVLVHVVVGNGMKEFMYAGQLIPCFEEAA 171

Qy        180 IVHINHGDVIRIPLFPVQLF---MPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHL 236

Db        172 PTRLNDCDAVRVPLYPPTLFGSLQADVD---SDELSLDKRS---SFVESRGLYVPAVSET 225

Qy        237 IHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARG 296

Db        226 LFYYVYTSWCQALRFSETKVLIEAALKQFVNDSQQSVKLAPHKKYFGYTSQK-------- 277

Qy        297 ARRNDVNSTSKPSPSGGFERRLASIMAADTAL------HAEVIFNTGIYEETPTDIKEWP 350

Db        278 -----LSSLEKD----------HLMLSDAVICELGFSFASVFLDSAYGASDSMVYSEWP 321

Qy        351 MFIGMEGTLPRLNALGSYTARVAGVIGAMVFSPNSAL------YLTEVEDSGMTEAKDGG 404

Db        322 VVVNATDHRDLIRALTELKLHLSTHISALLFSCNSILYHNRLVYLTSNKNASGT----GA 377

Qy        405 PGPSFNRFYQFAGPHLAANPQTDRDGH-VLSSQSTGSSNTEFSVDYLALICGFGAPLLAR 463

Db        378 SQEVLLKSIHFANGLTGLCEDTYNDARKLIKCSGVVAKDERYAPYHLSLICGTCPQLFSA 437

Qy        464 LLFYLERCDAGAFTGGHGDALKYVTGTFDSE------IPC--SLCEK------HTRPVCA 509

Db        438 FIWYLNRVSV-----------YNTGLTGSSTLSNHLIGCSSSLCGACGGTCCHT---CY 482

Qy        510 HTTVHRLRQRMPRFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEA---- 565

Db        483 NTAFVRVQTRLPQMPRLPKKEPSVVVMQSRFLNDVDVLGTFG----RRYSAESKEASLDA 538

Qy        566 ------------AKATMQDTYRATLERLFIDLEQERLLDRGAPCSSEGLSSVIVD-HPT 611

Db        539 KADEGSASTSNRTASSSVDRTHR--LNRILDYCKKMRLID-----SVTGEDTMTINGRSD 591

Qy        612 FRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPYSGAFCPITNFL 671

Db        592 FINLVSSLNKFVDDEAMSFVS---EVRMKSNRDEVLGATQAFNLDLNPFAVSFSPILAYE 648

Qy        672 VKRTHLAVVQDLALSQCHCVFYGQ----QVEGRNFRNQFQPVLRRRFVDLFNGGFISTRS 727

Db        649 YYRVIFAIIQNVALITATSYIVDNPLTTSLVSRWVTQHFQSI---------HGAFSTTSS 699

Qy        728 ITVTLSEGPVSAPNPTLGQDAPAGR--------------TFDGDLARVSVEVIRDIRVK 772

Db        700 -----RKGFLFIRNVKSSKNADHDRLPDFKLYARGTYSVISMEIKLSRLSVPSLLMFRVK 754

Qy        773 NRVVFSGNCTNLSEAARARLVGLASAYQRQE--KRVDMLHGALGFLLKQFHGLLFPRGMP 830

Db        755 NRPI--------SKASKGT---TAHVFFRREHVPKKNPVKGCLGFLLYKYHDKLFP---- 799

Qy        831 PNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRF-------TEEYAAINFINLPP 883

Db        800 --DCGFSCLQFWQKVCANALP----KNVNIGDMGEFNNFVKFVISVTADYNEHDLIDVPP 853

Qy        884 TCIGELAQFYMANLILKYCDHSQYL--INTLTSIITGARRPRDPSSVLHWIRKDVTSAAD 941

Db        854 DCMLNYLENRFHNKFLCFYGFKDYIGTLHGLTTRLTYQNHAQFP--YLLGESPNFASAAD 911

Qy        942 IETQAKALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQA 1001

Db        912 FALRLKDL--KATGVTAPLASTVTRESLMRTIFEQRSLVTVSFSIEKYAGVNNNKEIYQF 969

Qy       1002 GN---WSG------LN----GGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRET 1048

Db        970 GQIGYFSGNGVERSLNTNSIGGQD----YKFMR-QRCILATKLSDVLI-----KRSRRDN 1019

Qy       1049 TLSDQVRGIIVSGGAMVQL------------AIYATVVRAVGARAQHMAFDDWLSLTDD- 1095

Db       1020 VLFDE---DIIKNRVMAALDSENLDVDPELMAMY----EILSTREEIPERDDVLFFVDGC 1072
```

```
Qy       1096 EFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGETPTNLAFNFDSCEPSH 1155
             : :|  | |   :: :     :::   :   ::|| :   | |                 |
Db       1073 QAVADSLMEKFSRLQEMGVDDFSL---VNLQQVLDSRPECGGGGGEV-----------H 1117

Qy       1156 DTTSNVLNISGSNISGSTVPGLK-RPPED--DELFDLSGIPIKHGNI 1199
              | ::    ||  : |  || |  |  ||  |   :| | | :
Db       1118 DLSALFTAASGEAVGNSV--GLNARGGEHAFDE--DCGLLPAKRGRL 1160


earch completed: January 30, 2004, 13:13:04
ob time : 22 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:10:47 ; Search time 27 Seconds
(without alignments)
4284.849 Million cell updates/sec

Title: US-09-769-699-2
Perfect score: 6294
Sequence: 1 MENTQKTVTVPTGPLGYVYA.....DELFDLSGIPIKHGNITMEM 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6294	100.0	1204	1	DNBE29	DNA-binding protei
2	3620.5	57.5	1209	1	DNBEC4	DNA-binding protei
3	3583	56.9	1208	2	T42574	DNA-binding protei
4	3054	48.5	1196	1	DNBEKS	DNA-binding protei
5	3049	48.4	1196	1	DNBEV1	major DNA-binding
6	3047	48.4	1196	1	DNBEHF	DNA-binding protei
7	3003	47.7	1197	1	A48350	DNA-binding protei
8	2933	46.6	1186	1	DNBEBG	DNA-binding protei
9	1094.5	17.4	375	2	JQ0846	DNA-binding protei
10	669	10.6	1128	1	QQBE47	DNA-binding protei
11	633	10.1	1128	2	T42922	major single-stran
12	618	9.8	1145	2	S55600	single-stranded DN
13	604	9.6	1128	1	DNBEM1	DNA-binding protei
14	596	9.5	1127	2	T03105	major single-stran
15	490.5	7.8	1132	2	T44001	major DNA binding
16	483	7.7	1131	2	T41943	major DNA binding
17	479.5	7.6	1160	2	A36256	DNA-binding protei
18	477	7.6	1235	1	QQBEW4	DNA-binding protei
19	412	6.5	1191	1	A44051	DNA-binding protei
20	237.5	3.8	483	2	S69894	major DNA-binding
21	137	2.2	1504	2	T17426	FK506 polyketide s
22	120.5	1.9	781	2	AI3014	methyl-accepting c
23	120.5	1.9	788	2	F98269	hypothetical prote
24	119.5	1.9	701	2	AI1501	internalin protein
25	117.5	1.9	358	2	H83380	O6-methylguanine-D
26	116.5	1.9	3335	2	H81702	adherence factor T
27	115.5	1.8	7576	2	T17428	FK506 polyketide s
28	115	1.8	823	2	B35963	protein-tyrosine k
29	115	1.8	2505	1	XYRTFA	enoyl-[acyl-carrie
30	114	1.8	1215	2	I52882	autoantigen - huma
31	114	1.8	1941	2	T23979	hypothetical prote
32	114	1.8	1943	2	T23986	hypothetical prote
33	113.5	1.8	1886	2	S04921	nuclear pore prote
34	113	1.8	1682	1	C70588	probable mbtE prot
35	112	1.8	735	2	F87355	GGDEF family prote
36	111	1.8	1699	2	T14074	complement compone
37	111	1.8	3255	2	G81702	adherence factor T
38	110.5	1.8	878	2	A87385	TonB-dependent rec
39	110.5	1.8	2609	2	T40399	probable transport
40	110.5	1.8	3069	2	H70656	fatty-acid synthas
41	110	1.7	824	2	S24108	protein-tyrosine k
42	110	1.7	1039	2	AI0122	probable integral
43	109.5	1.7	977	2	G72526	hypothetical prote
44	109.5	1.7	1501	2	T45623	hypothetical prote
45	109.5	1.7	1762	2	T03222	probable polyketid

ALIGNMENTS

RESULT 1
DNBE29
DNA-binding protein - human herpesvirus 3
C;Species: human herpesvirus 3, varicella-zoster virus
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C;Accession: C27214
R;Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A;Title: The complete DNA sequence of varicella-zoster virus.
A;Reference number: A27345; MUID:86306657; PMID:3018124
A;Accession: C27214
A;Molecule type: DNA
A;Residues: 1-1204 <DAV>
A;Cross-references: EMBL:X04370; NID:g59989; PIDN:CAA27912.1; PID:g60018
C;Genetics:
A;Gene: 29
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding

Query Match 100.0%; Score 6294; DB 1; Length 1204;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120

Qy        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180

Qy        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240

Qy        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300

Qy        301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLP 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLP 360

Qy        361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHL 420

Qy        421 AANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480
```

09/769,699
Search Notes
2-2-04

```
Db        421 AANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480

Qy        481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNSQ 540

Db        481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNSQ 540

Qy        541 YSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSE 600

Db        541 YSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSE 600

Qy        601 GLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660

Db        601 GLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660

Qy        661 SGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFNG 720

Db        661 SGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFNG 720

Qy        721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGN 780

Db        721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGN 780

Qy        781 CTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW 840

Db        781 CTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW 840

Qy        841 FWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900

Db        841 FWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900

Qy        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960

Db        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960

Qy        961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFD 1020

Db        961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFD 1020

Qy       1021 RTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGAR 1080

Db       1021 RTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGAR 1080

Qy       1081 AQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGET 1140

Db       1081 AQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGET 1140

Qy       1141 PTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHGNIT 1200

Db       1141 PTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHGNIT 1200

Qy       1201 MEM 1203

Db       1201 MEM 1203
```

```
ESULT 2
)NBEC4
)NA-binding protein - equine herpesvirus 1 (strain Ab4p)
:;Species: equine herpesvirus 1
;Note: host Equus caballus (domestic horse)
:;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
:;Accession: E36798
:;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
ubmitted to GenBank, March 1992
;Description: The DNA sequence of equine herpesvirus-1.
;Reference number: A36805
;Accession: E36798
.;Molecule type: DNA
.;Residues: 1-1209 <TEL>
;Cross-references: GB:M86664; NID:g330791; PIDN:AAB02466.1; PID:g330823
;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
irology 189, 304-316, 1992
```

```
A;Title: The DNA sequence of equine herpesvirus-1.
A;Reference number: A41831; MUID:92295566; PMID:1318606
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 31
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding

  Query Match             57.5%;  Score 3620.5;  DB 1;  Length 1209;
  Best Local Similarity   57.1%;  Pred. No. 1e-266;
  Matches  698;  Conservative  191;  Mismatches  299;  Indels   35;  Gaps   14;

Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60

Db          1 MESAPKTVSLPVSPLGYVYARQKASLQTGTVSLTAARSVDSDLAVLPVIRGLTVEQTFTT 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120

Db         61 NVAVVAGSKTTGLGGTGITLKLTPSHFNPNAFVFYGGSVIGASSNAPNLTRACEAARRRF 120

Qy        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180

Db        121 GFSAFSSPPVENAVETSGEEICASLNLSPETTALYLVVTESFKEMVYVCNTFLHYGGTST 180

Qy        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240

Db        181 VTIDGQDAMKIPIYPVQLYMPDVNRLASEPFNAKHRSIGDEFVYSRPFFNSDLCRLLHGY 240

Qy        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSS--SSGTTTARGAR 298

Db        241 VLGPAAVALRVRNLDGVARGAAHLALDENHEGSVLPQDVTFTLFDSTQGNAGKGSGRAQR 300

Qy        299 RNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGT 358

Db        301 QGD-GSGSKNSASSGIERRLASVMAADTALSVDSIMGAGIYDTELPSVEDWPVLSSGDDT 359

Qy        359 LPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGP 418

Db        360 -ESLEALGAYAARLSGLVGAMVFSANSVLYMTEVDDGGPADGKDGS-NPSYHRFYLIAAP 417

Qy        419 HLAANPQTDRDGHVL------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCD 472

Db        418 YVAGNPQTDKDGRVLPHTADQQAAPINGSNQEFSLDYLALACGFCPQILARLLFYLERCD 477

Qy        473 AGAFTG-GHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPI 531

Db        478 AGTFGGRNETDALRYLANTLESDVPCGLCNQATRPACAHTTLHRLRQRLPRFGAPVRAPI 537

Qy        532 GVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLL 591

Db        538 GIFGTMNSAYSDCDVLGNYASYGALKRPND-NEAPKSIMQDTYRATMERLVNELEQAKLI 596

Qy        592 DR-----GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGL 646

Db        597 DKETLAQASPCSAP--TSVVHDQASFIGLLSNIKDTIEGAAEQFMRTLVEARDFKIREGL 654

Qy        647 SEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQF 706

Db        655 ADANHTMSISLDPYSSSFCPVTSFLARRTVFAVLQDLVLSQCHCLFYGQSVEGRNFRNQF 714

Qy        707 QPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766

Db        715 QPVLRRRFLDMLNGGFITAKTVTVTVSDSGVLAPDLTRPASEPPTKDYDGDMARVSMEVL 774

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826

Db        775 RDLRVKNRVLFSNGGANMSEAARARVAGMASAYRRPDKGSNILNGAVGFLVKQYHGVLFP 834

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886

Db        835 RGHPPGIDTPNPQWFWTLLQRNQMPARLLSKEDIETITAIKRFSDEYSAINFINLTPNNI 894
```

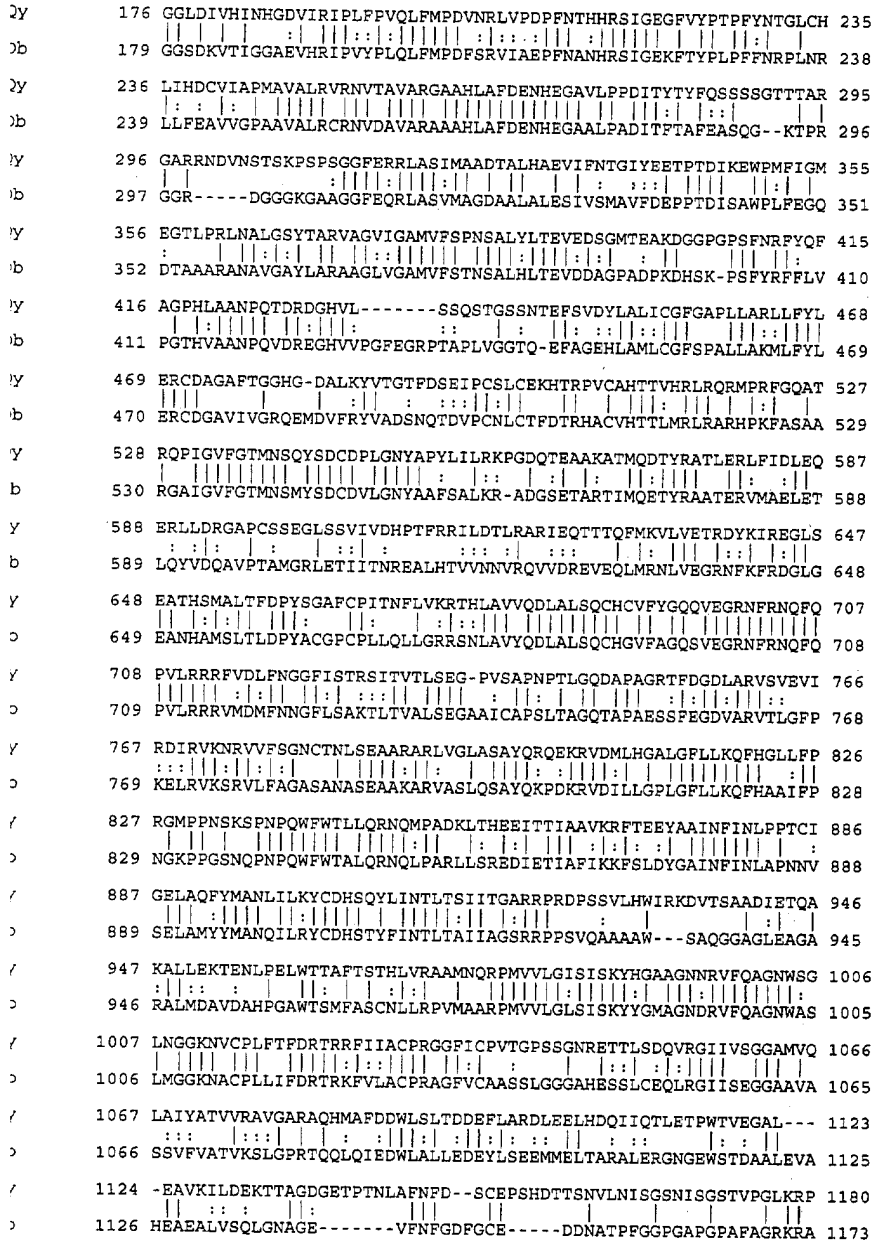
```
Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
Db        895 GELAQFYFANLVLKYCDHSQYFINGLTAIVVGSRRPRDPAAVLAWIDRTINGAADVEPAA 954

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
Db        955 QEVLQRLGSNPAAWTGTFTSTNMVRYVMDQRPMVVIGLSISKYNGSAGNNRVFQAGNWNG 1014

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
Db       1015 LNGGKNVCPLMAFDRTRRFVLACPRVGFTCEAGGFGTGVRENTLSEQVRGIVSEGGPMVQ 1074

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAV 1126
Db       1075 TAVFAAVLHALGARTQHLAVDDWIGLVDDEFLAASLDALNATVVDQF-GEWSVEAAQELV 1133

Qy       1127 KILDEKTTAG---DGETPTNLAFNFDSC---EPSHDTTSNVLNISGSNISGSTVPGLKRP 1180
Db       1134 KNMEAQTTAGAVAAGEG----AFDFGACVGDTPQQSTSA----FNGGLAMAAAPAGQKRS 1185

Qy       1181 PEDDELFDLSGIPIKHGNITMEM 1203
Db       1186 LPDDILFDMGAPPEKKSGLTFDM 1208
```

ESULT 3
'42574
NA-binding protein - equine herpesvirus 4 (strain NS80567)
;Species: equine herpesvirus 4
;Variety: strain NS80567
;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
;Accession: T42574
;Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
. Gen. Virol. 79, 1197-1203, 1998
;Title: The DNA sequence of equine herpesvirus-4.
;Reference number: Z22173; MUID:98264497; PMID:9603335
;Accession: T42574
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-1208 <TEL>
;Cross-references: EMBL:AF030027; NID:g2605950; PIDN:AAC59547.1; PID:g2605975
;Experimental source: strain NS80567
;Genetics:
;Gene: 31
;Superfamily: herpesvirus DNA-binding protein
;Keywords: DNA binding

```
 Query Match             56.9%;  Score 3583;  DB 2;  Length 1208;
 Best Local Similarity   56.4%;  Pred. No. 7.4e-264;
 Matches  689;  Conservative  194;  Mismatches  305;  Indels   34;  Gaps   13;

          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
          1 MESAPKTVSLPVSPLGYVYAIQNTFMETEALTLMAARSIDSDLAVLPVIRGLTVEQTFTT 60

         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
         61 NVAVVAGSKTTGLGGAGITLKLTPSHFTPNAFVFYGGSVFGASSKAPNLTRACELARRRF 120

        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
        121 GFSPFSSPPVDNAVETSGEEICASLNLSPETTTLYLVVTETFKEMVYMCNTFLHYGGTST 180

        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
        181 VTIHGQEAVKIPIYPVQLYMPDVNRLAAEPFNSKHRSIGDEFVYSKPFFNSDLCRLLHGY 240

        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSS--SSGTTTARGAR 298
        241 VLGPAAVALRVRNLDGVARGAAHLALDENHEGSVLPQDVTFTLFDSAQGTSGKGSGRTQR 300

        299 RNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGT 358
```

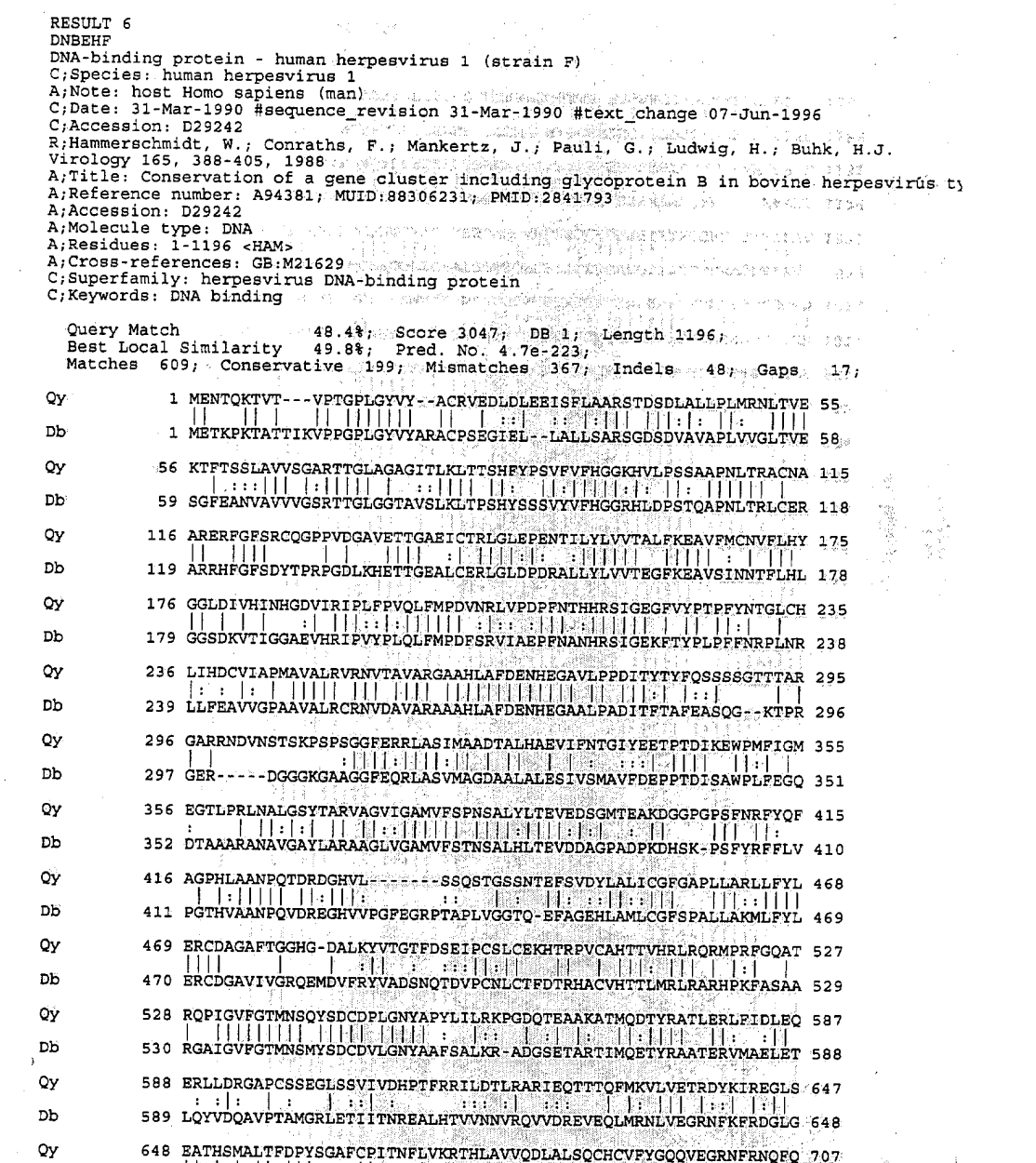
```
Db        301 QGD-GSGLKNGSSSGIERRLASIMAADTALSVDSIMGAGVYDTELPSVEDLPI-LSVGDD 358

Qy        359 LPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGP 418
Db        359 RERLEALGAYASRLSGLVGAMVFSANSVLYMTEVDDGGPADGKDAS-NPSYHRFYLIAAP 417

Qy        419 HLAANPQTDRDGHVLS------SQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCD 472
Db        418 YVAGNPQTDKDGRVLQHTADQPAAPINGSNQEFSLDYLALACGFCPQLLARILFYLERCD 477

Qy        473 AGAFTG-GHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPI 531
Db        478 AGTFGGRNETDALRYLANTLESEVPCGLCTPATRPACAHTTLHRLRQRLPRFGTPVRAPI 537

Qy        532 GVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLL 591
Db        538 GIFGTMNSTYSDCDVLGNYASYGALKRPND-NEAPKSIMQDTYRATMERLVNDLEQAKLI 596

Qy        592 DRGA-----PCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGL 646
Db        597 DKEALAHAGTCSAS--TGVVKDQASFINLLSTIKDITEGAAEQFMRTLVEVRDFKIREGL 654

Qy        647 SEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQF 706
Db        655 ADANHTMSISLDPYSSSFCPVTSFLSRRTIFAVLQDLVLSQCHCLFYGQSVEGRNFRNQF 714

Qy        707 QPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766
Db        715 QPVLRRRFLDMLNGGFITAKTVTVTVSDSGVTAPNLTLPSSEPPTKDYDGDMARVSMEVL 774

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826
Db        775 RDLRIKNRVLFSNGGANMSEAARARVAGMASAYRRPEKGSNILNGAVGFLVKQFHKVLFP 834

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886
Db        835 RGHPPGIDTPNPQWFWTLLQRNQMPARLLSKEDIETITAIKRFSHEYSAINFINLTPNNI 894

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
Db        895 GELAQFYFANLVLKYCDHSQYFINGLTAIVVGSRRPRDPAAVLAWINRTINGASDVEPAA 954

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
Db        955 QEVLQQLGSNPAAWTGTFASTNMVRYVMDQRPMVVIGLSISKYNGSAGNNRVFQAGNWNG 1014

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
Db       1015 LNGGKNVCPLMAFDRTRRFVLACPRVGFTCEAGGFGMGARENTLSEQIRSIVSDGGPMVQ 1074

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAV 1126
Db       1075 TAVFSVVLTALGARTQHLAVDDWIGLVDDEFLAASLDALNAAVVDQF-GEWSVEAAQDMI 1133

Qy       1127 KILDEKT-----TAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPP 1181
Db       1134 RTMDAQTNMGVVSTGDG------AFDFGACVGDANQSSTTFNMGPA--SSSAPAGQKRFH 1185

Qy       1182 EDDELFDLSGIPIKHGNITMEM 1203
Db       1186 PDDILFDMGAPPEKKSGLTFDM 1207
```

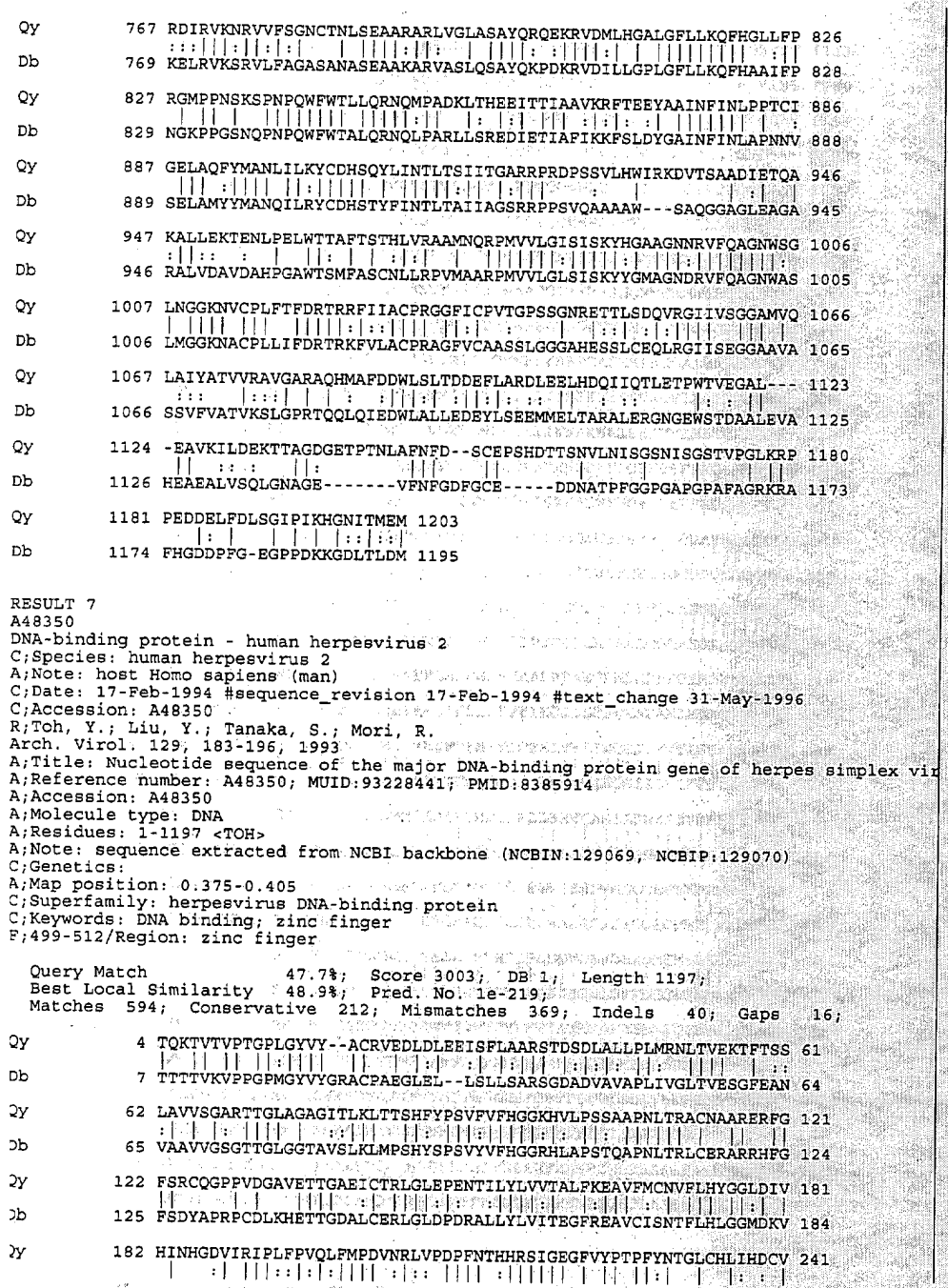
RESULT 4
DNBEKS
DNA-binding protein - human herpesvirus 1 (strain KOS1.1)
C;Species: human herpesvirus 1
A;Note: host Homo sapiens (man)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
C;Accession: A28601
R;Gao, M.; Bouchey, J.; Curtin, K.; Knipe, D.M.
Virology 163, 319-329, 1988
A;Title: Genetic identification of a portion of the herpes simplex virus ICP8 protein rec



RESULT 6
DNBEHF
DNA-binding protein - human herpesvirus 1 (strain F)
C;Species: human herpesvirus 1
A;Note: host Homo sapiens (man)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 07-Jun-1996
C;Accession: D29242
R;Hammerschmidt, W.; Conraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.
Virology 165, 388-405, 1988
A;Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus t
A;Reference number: A94381; MUID:88306231; PMID:2841793
A;Accession: D29242
A;Molecule type: DNA
A;Residues: 1-1196 <HAM>
A;Cross-references: GB:M21629
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding

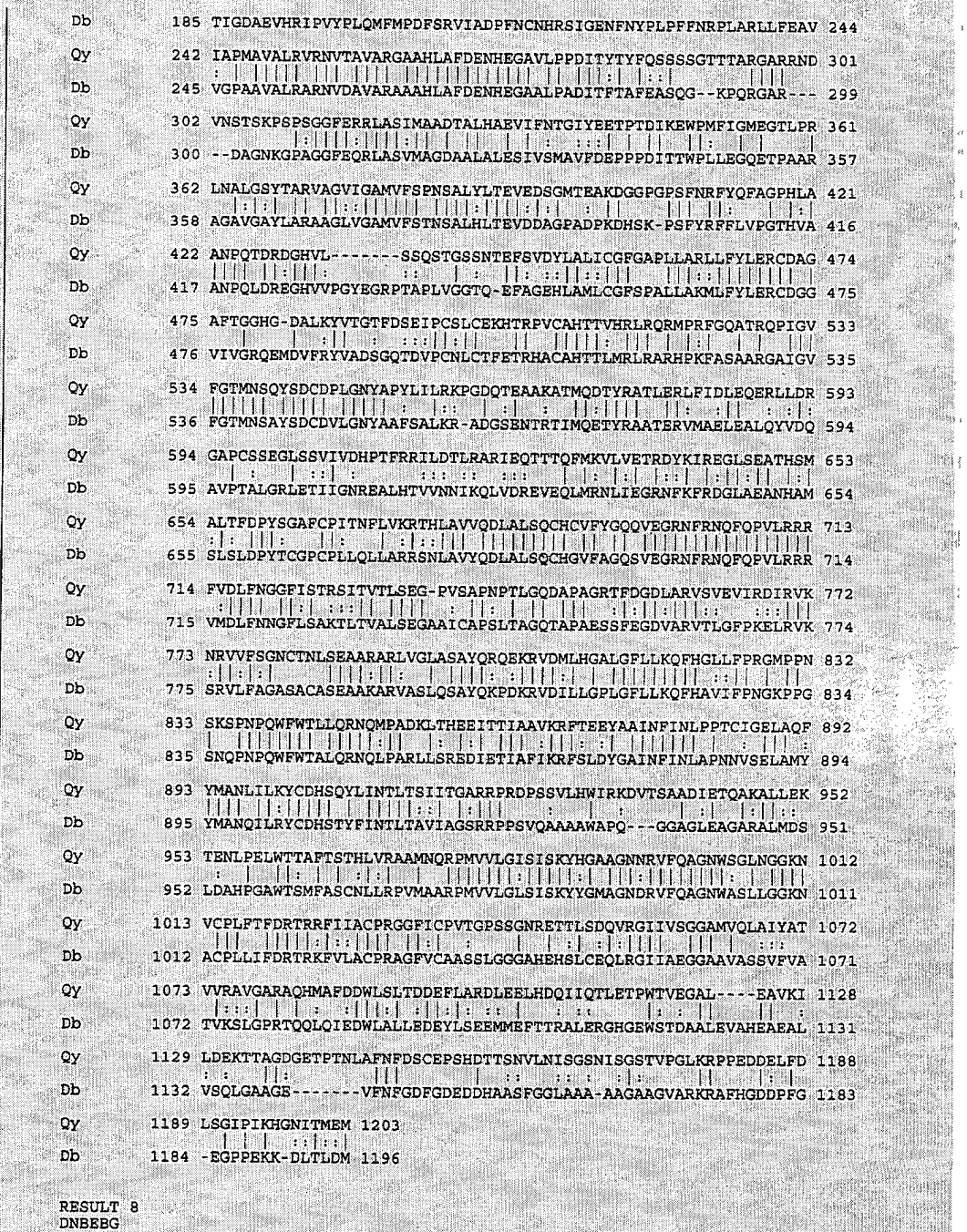
Query Match 48.4%; Score 3047; DB 1; Length 1196;
Best Local Similarity 49.8%; Pred. No. 4.7e-223;
Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;





RESULT 7
A48350
DNA-binding protein - human herpesvirus 2
C;Species: human herpesvirus 2
A;Note: host Homo sapiens (man)
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 31-May-1996
C;Accession: A48350
R;Toh, Y.; Liu, Y.; Tanaka, S.; Mori, R.
Arch. Virol. 129, 183-196, 1993
A;Title: Nucleotide sequence of the major DNA-binding protein gene of herpes simplex vir
A;Reference number: A48350; MUID:93228441; PMID:8385914
A;Accession: A48350
A;Molecule type: DNA
A;Residues: 1-1197 <TOH>
A;Note: sequence extracted from NCBI backbone (NCBIN:129069, NCBIP:129070)
C;Genetics:
A;Map position: 0.375-0.405
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding; zinc finger
F;499-512/Region: zinc finger

Query Match 47.7%; Score 3003; DB 1; Length 1197;
Best Local Similarity 48.9%; Pred. No. 1e-219;
Matches 594; Conservative 212; Mismatches 369; Indels 40; Gaps 16;



RESULT 8
DNBEBG

DNA-binding protein - bovine herpesvirus 2 (strain BMV)
C;Species: bovine herpesvirus 2
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 07-Jun-1996
C;Accession: A29242
R;Hammerschmidt, W.; Conraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.
Virology 165, 388-405, 1988
A;Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus t
A;Reference number: A94381; MUID:88306231; PMID:2841793
A;Accession: A29242
A;Molecule type: DNA
A;Residues: 1-1186 <HAM>
A;Cross-references: GB:M21628
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding

Query Match 46.6%; Score 2933; DB 1; Length 1186;
Best Local Similarity 48.7%; Pred. No. 2.2e-214;
Matches 585; Conservative 188; Mismatches 373; Indels 56; Gaps 11;

```
Qy          1 MENTQKTVT---VPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKT 57
              |||  |||  |    |  ||||||| |:     | |:: |:||| ||| |:|||:  ||||
Db          1 MENKQKTATTVKVSPGPLGYVYARRLPPEGLTELALLSARSADSDTAVLPLIAGLTVESG 60

Qy         58 FTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAAR 117
              |  ::||| |:|||: | |::|||  ||: || :|||||:|: ||||||||:  |: ||
Db         61 FDVNVAVVVGSRTTGVGGTGVSLKLMPSHYAPSAYVFHGGRHLAPSSAAPNLSLLCDRAR 120

Qy        118 ERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGG 177
              :||||  :  | :  ||||  :|  ||| |  ::||:|:   ||||:: |  || ||
Db        121 VQFGFSSFKPKPCEAEGETTGEALCEHLGLNPNESLLYMVIAEGFKEAVYISNTILHMGG 180

Qy        178 LDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLI 237
              :  | |   :| |||::|:|:||||  | | ||||  ||:||| | || ||:|  |  |:
Db        181 VGTVTIAGEEVRRIPIYPLQMFMPDYCRAVADPFNDRHRAIGEYFAYPLPFFNAKLASLL 240

Qy        238 HDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGA 297
              : | ||||| ||| |||| ||||||||||||| || |||:| |  : |     :
Db        241 FGAAVGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFDPTGS-----KAG 295

Qy        298 RRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEG 357
              ||          ||||:||||:|| | ||  | | :  ::|| ||||  |||     |
Db        296 HRNPREC------GGGFEQRLASVMAGDAALALESIMSMAVFEEPPTDIGTWPMLTCQES 349

Qy        358 TLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAG 417
              |  |  ::|:|  | ||::|||||| ||||:||||:|:|  : ||    ||| ||:   |
Db        350 TAARAASIGAYLGRAAGLVGAMVFSSNSALHLTEVDDAGPADPKD-PTKPSFYRFFLVPG 408

Qy        418 PHLAANPQTDRDGHVLSSQS------TGSSNTEFSVDYLALICGFGAPLLARLLFYLERC 471
              ::|||||| |||| |::          | ||: ::|| :|||    |||::|:|||||
Db        409 TYVAANPQLDRDGRVVAGHEGRPIVPIVGGNHEFTCEHLATLCGFSPELLAKMLYYLERC 468

Qy        472 DAGAFTG-GHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQP 530
              | |   |   |  |||: :  :::|| ||    |  |||||: ||| | |:|   ||
Db        469 DGGVILGRPEMDTFKYVSDSAHTDVPCCLCSLDNRHSCAHTTLLRLRARHPKFTSTTRGA 528

Qy        531 IGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERL 590
              ||:|| ||| ||||| ||||| :  :::  |  | |:| ||:|||: :||:  :||
Db        529 IGIFGVMNSAYSDCDVLGNYASFSAIKRM-DVQETARAIMQETYRSAVERVMAELENLNY 587

Qy        591 LDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEAT 650
              :|   | |   | |:|      : ::  ::  ::     | |: ||| | :: || | ||
Db        588 IDAAVPTSPAKLESIITGREALQTVVSNVKQVVDGEVAQLMRALVEGRGFRFREALGEAN 647

Qy        651 HSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVL 710
              |:|:|| ||::   ||:   | :|::||| |||||||| || || |||||||:||||||
Db        648 HAMSLTLDPHASVPCPLLQMLGRRSNLAVYQDLALSQCHGVFEGQAVEGRNFRSQFQPVL 707

Qy        711 RRRFVDLFNGGFISTRSITVTLSEGP-VSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDI 769
              ||| :|:|| ||:| |::|| |::|  :|||    || | |   |:||:|||::   ::|
Db        708 RRRVLDMFNNGFLSARTLTVALTDGACISAPGLVSGQHAAAESGFEGDVARVNLGFPKEI 767

Qy        770 RVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGM 829
              |||:||:|:|     |||||||: || |||| :|||||:| | |||:|||||  ||| |

Db        768 RVKSRVLFAGAGPAASEAARARIAGLQSAYQKSDKRVDILLGPLGFMLKQFHATLFPNGK 827

Qy        830 PPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGEL 889
              || | :|||||||| ||||:||   |:|:|: || |||| || | ||:|||   | ||
Db        828 PPGSDNPNPQWFWTALQRNQLPARLLSREDISLIAFVKRFSVEYGAGNFVNLPPNNISEL 887

Qy        890 AQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKAL 949
              | :|||| |||||||| | |||||::| |:||| :  :  |  :  |    ::| ||:::
Db        888 AMYYMANQILKYCDHSTYFINTLTALIAGSRRPPNAQAAAAWAPRGGT---ELEAQARSV 944

Qy        950 LEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNG 1009
              :   : |  ||| | | ::|:|  |  |||||||:|||||:| |||:|||||| : | |
Db        945 VANPGDHPGAWTTMFASCNLLRPVMATRPMVVLGLSISKYYGMAGNDRVFQAGNLANLLG 1004

Qy       1010 GKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAI 1069
              ||| |||  |||||:|:|||| ||:|     ||  |::| :|:||  ||| |   :
Db       1005 GKNACPLLIFDRTRKFVIACPRAGFVCAAVSAGSGAHESSLCEQLRAIIAEGGATVASDV 1064

Qy       1070 YATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKIL 1129
              :|   :::||| | :  :|||:| :||:|: :: ||  : ::      |::: ||:  :
Db       1065 FAAAAKSLGARVQQLQIEDWLALLEDEYLSEEMMELAGRALERGGGEWSLDAALDVAREA 1124

Qy       1130 DEKTT--------------AGDGETPTNLAFNFDS---------------CEPSHDTTSN 1160
              :  |                | ||     || :  |                  :| | :
Db       1125 EAMVTRHVDAEETFDFGAFAEDGPADAGLAVHLQSRRRPLACSDLFGDAPAEKRNDLTLD 1184

Qy       1161 VL 1162
              :|
Db       1185 ML 1186
```

RESULT 9
JQ0846
DNA-binding protein - equine herpesvirus 1 (fragment)
C;Species: equine herpesvirus 1
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Feb-1997
C;Accession: JQ0846
R;Bell, C.W.; Whalley, J.M.
submitted to JIPID, January 1991
A;Reference number: JQ0846
A;Accession: JQ0846
A;Molecule type: DNA
A;Residues: 1-375 <BEL>
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding; nucleus

Query Match 17.4%; Score 1094.5; DB 2; Length 375;
Best Local Similarity 55.5%; Pred. No. 3.1e-75;
Matches 212; Conservative 57; Mismatches 98; Indels 15; Gaps 5;

```
Qy        828 GMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIG 887
              | ||   :|||||||||||||||||  |: :|: || |:|||:|:|||||| | :||
Db          2 GHPPGIDTPNPQWFWTLLQRNQMPARLLSKEDIETITAIKRFSDEYSAINFINLTPNNIG 61

Qy        888 ELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAK 947
              ||||||  |||:|||||||| || ||:|: :|||||::||  : :  :|||:|  |:
Db         62 ELAQFYFANLVLKYCDHSQYFINGLTAIVVGSRRPRDPAAVLAWIDRTINGAADVEPAAQ 121

Qy        948 ALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGL 1007
              :|::  :  |  ||  |||||:||  |:|||||:|:|||||:|:|||||||:|| ||
Db        122 EVLQRLGSNPAAWTGTFTSTNMVRYVMDQRPMVVIGLSISKYNGSAGNNRVFQAGNWNGL 181

Qy       1008 NGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQL 1067
              |||||||||  |||||||:|||| ||  |  :| || |||:|||||:  || |||
Db        182 NGGKNVCPLMAFDRTRRFVLACPRVGFTCEAGGFGTGVRENTLSEQVRGIVSEGGPMVQT 241

Qy       1068 AIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVK 1127
              |::| |: |:||| ||:|:|||||:| |||||  |:  :  :|||| | | ||| |||
Db        242 AVFAAVLHALGARTQHLAVDDWIGLVDDEFLAASLDALNATVVDQF-GEWSVEAAQELVK 300

Qy       1128 ILDEKTTAG---DGETPTNLAFNFDSC---EPSHDTTSNVLNISGSNISGSTVPGLKRPP 1181
              :: :||||    ||     ||:| :|    |  |::   :|    :  | ||
Db        301 NMEAQTTAGAVAAGEG----AFDFGACVGDTPQQSTSA----FNGGLAMAAAPAGQKRSL 352
```

```
Qy       1182 EDDELFDLSGIPIKHGNITMEM 1203
Db        353 PDDILFDMGAPPEKKSGLTFDM 374

RESULT 10
QQBE47
DNA-binding protein - human herpesvirus 4 (strain B95-8)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999
C;Accession: A43045; A03791; S33057
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A;Reference number: A93065; MUID:85035713; PMID:6092825
A;Accession: A43045
A;Molecule type: DNA
A;Residues: 1-1128 <BAN>
A;Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24808.1; PID:g1334916
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667; PMID:6087149
A;Contents: annotation; protein coding region
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding

  Query Match             10.6%;  Score 669;  DB 1;  Length 1128;
  Best Local Similarity   23.2%;  Pred. No. 4.8e-42;
  Matches  292;  Conservative  186;  Mismatches  537;  Indels  246;  Gaps   45;

Qy         13 GPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGARTTG 72
Db         18 GPCGYIYFYPLATYPLREVATLGTGYAGHRCLTVPLLCGITVEPGFSIN---VKALHRRP 74

Qy         73 LAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFS------RCQ 126
Db         75 DPNCGL---LRATSYHRDIYVFHNAHMVPPIFEGPGLEALCGETREVFGYDAYSALPRES 131

Qy        127 GPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINHG 186
Db        132 SKPGDFFPE----------GLDPSAYLGAVAITEAFKERLYSGNLVAIPSLKQEVAVGQS 181

Qy        187 DVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMA 246
Db        182 ASVRVPLYDKEVFPEGVPQL------------------RQFYNSDLSRCMHEALYTGLA 222

Qy        247 VALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTS 306
Db        223 QALRVRRVGKL--------------------VELLEKQSLQDQAKVAKVAPLKEFPAST 261

Qy        307 KPSPSGGFERRLASIMAADT-ALHAEVIFNTGIYE---ETPTDIK--EWPMFIGMEGTLP 360
Db        262 ISHPDSG------ALMIVDSAACELAVSYAPAMLEASHETPASLNYDSWPLFADCEGPEA 315

Qy        361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAG--- 417
Db        316 RVAALHRYNASLAPHVSTQIFATNSVLYV-----SGVSKSTGQGKESLFNSFYMTHGLGT 370

Qy        418 -----------PHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLF 466
Db        371 LQEGTWDPCRRPCFSGWGGPDVTG------TNGPGN--YAVEHLVYAASFSPNLLARYAY 422

Qy        467 YLERCDAGAFTGGHGDAL-------KYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQR 519
Db        423 YLQFCQ------GQKSSLTPVPETGSYVAGAAASPM-CSLCEGRAPAVCLNTLFFRLRDR 475

Qy        520 MPRFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLIL----RKPGDQTEAAKATMQDTYR 575
Db        476 FPPVMSTQRRDPYVISGASGSYNETDFLGNFLNFIDKEDDGQRPDDEPRYTYWQLNQNLL 535

Qy        576 ATLERLFIDLEQERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLV 635
Db        536 ERLSRLGIDAEGKLEKEPHGP------------RDFVKMFKDVDAAVDAEVVQFMNSMA 582

Qy        636 ETRDYKIREGLSEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQ 695
Db        583 K-NNITYKDLVKSCYHVMQYSCNPFAQPACPIFTQLFYRSLLTILQDISLPICMC----- 636

Qy        696 QVEGRNFRNQFQPVLRRRFVDLFNGGF----ISTRSITV-----TLSEGPVSAPNPTLG- 745
Db        637 ------YEND-NPGLGQSPPEWLKGHYQTLCTNFRSLAIDKGVLTAKEAKVVHGEPTCDL 689

Qy        746 -------QDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASA 798
Db        690 PDLDAALQGRVYGRRLPVRMSKVLMLCPRNIKIKNRVVFTGENAALQNS-------FIKS 742

Qy        799 YQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQRNQMPADK-LTH 857
Db        743 TTRRENYI--INGPYMKFLNTYHKTLF-----PDTKLSSLYLWHNFSRRRSVPVPSGASA 795

Qy        858 EEITTIAA-VKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSII 916
Db        796 EEYSDLALFVDGGSRAHEESNVIDVVPGNLVTYAKQRLNNAILKACGQTQFYISLIQGLV 855

Qy        917 --TGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAM 974
Db        856 PRTQSVPARDYPHVLG--TRAVESAA-------AYAEATSSLTAT-TVVCAATDCLSQVC 905

Qy        975 NQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGG- 1033
Db        906 KARPVVTLPVTINKYTGVNGNNQIFQAGN-LGYFMGRGV------DRN---LLQAPGAGL 955

Qy       1034 --------------FICPVTGPSSGNR-ETTLSDQVRGIIVSGGAMV-----QLAIYATV 1073
Db        956 RKQAGGSSMRKKFVFATPTLGLTVKRRTQAATTYEIENIRAGLEAIISQKQEEDCVFDVV 1015

Qy       1074 VRAVGARAQHMAFDDWLSLT--DDEFLARDLEELHDQIIQTLET------PWTVEGALEA 1125
Db       1016 CNLVDAMGEACA-----SLTRDDAEYLLGRFSVLADSVLETLATIASSGIEWTAEAARDF 1070

Qy       1126 VKILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISG--------STVPGL 1177
Db       1071 L----EGVWGGPGAAQDNFISVAEPVSTASQASAGLLLGGGGQGSGGRRKRRLATVLPGL 1126

Qy       1178 K 1178
Db       1127 E 1127

RESULT 11
T42922
major single-stranded DNA binding protein - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C;Accession: T42922
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274
A;Accession: T42922
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1128 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95531.1
A;Experimental source: strain 73
C;Superfamily: herpesvirus DNA-binding protein

  Query Match             10.1%;  Score 633;  DB 2;  Length 1128;
  Best Local Similarity   22.4%;  Pred. No. 2.6e-39;
  Matches  274;  Conservative  194;  Mismatches  514;  Indels  240;  Gaps   41;

Qy         11 PTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGART 70
```

```
Db        18 PVGACGYIYVYPKEGPPFQEASLLGNRNVGADAMSLPLLSGLTVEANFSFNVKAVHKKID 77
Qy        71 TGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPPV 130
Db        78 M------TTLSVRVSAYHREAIVFFNTNFFTPIFPGPGLDSLCSDARNLFGYTSFT-PRH 130
Qy       131 DGA----VETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINHG 186
Db       131 DLKDLIDIKDLYAPFYTK-----DSCFMSVVVTESFKERLYFGNLVPLISQGQKVQINGR 185
Qy       187 DVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMA 246
Db       186 EAVKIPLYDEDLFSKSHEHL------------------PRFYIPSVSKYLHDSLFTSIA 226
Qy       247 VALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTS 306
Db       227 QALRIRDVESVIRAIEKQSIHDQYKLAKV------------------------VNSK- 259
Qy       307 KPSPSGGFERRLASIMAADTA-------LHAEVIFNTGI-YEETP------TDIKEWPMF 352
Db       260 --------EFALQAVKCQDASAFMVIDCIAAELAISYGLSFLETPQEPCALLDYTSWPIF 311
Qy       353 IGMEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRF 412
Db       312 DTTETEEGRIKAIQDWNAMMSVHVYTHLFSTNSVLYLTKINKQTQTNKSDQN---IYNSY 368
Qy       413 YQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCD 472
Db       369 FMQHGLSYAAEATQKENGDPAFSGAVKFNGGSYTLFHLALASSFSPHLLARNCYYMQFCQ 428
Qy       473 AGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIG 532
Db       429 HQKSTTNSNYSVAQYVGTVAVSELCELCQGKCPAACIHTLFYRLKDRFPPVLSSQRRDPY 488
Qy       533 VFGTMNSQYSDCDPLGNYAPYLILRKPGD----QTEAAKATMQDTYRATLERLFIDLEQE 588
Db       489 VVTGVSGQYNDLDMLGNFATF---REKEDDMVQNTECEKYTYWQLIQNVVEKL------- 538
Qy       589 RLLDRGAPCSSEGL--SSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGL 646
Db       539 -----ATIGITEGTLGSDLITDIQSFLKTFREIDNVVDNEVIKFINCLVK-NNINFRETI 592
Qy       647 SEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQ----------- 695
Db       593 KGVHHVLHYCCNVFWQAPCAMFLNLFYKSVLAIIQDICLPV--AMIYEQDNPSMGMMPSE 650
Qy       696 ------QVEGRNFRN----------QFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS 738
Db       651 WLKMHYQTIWTNFKSSCIDKGVLTGAEHKVVHGDMFCDFINVD---------------S 694
Qy       739 APNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASA 798
Db       695 ALN---GQIVPV--KMQVRLAKALITVPKTIKIKNRIVFSN--SSMTEAIQAGFI----- 742
Qy       799 YQRQEKRVD--MLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFW-TLLQRNQMPA-DK 854
Db       743 --RSSTKKDSYIVTGPYMKFLNSLHKVIFP------DAKISALYLWHTFSQKRQIPVLPG 794
Qy       855 LTHEEITTIA-AVKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINTL- 912
Db       795 ISKENIIELANYIEAGSKMHDDMNVLDIIPTTLLAYAKVRLNNTILRTCGQTQFYATTLQ 854
Qy       913 -----TSIITGARRPRDPSSVLHWIRKDVTSAADIETQAK---ALLEKTENLPELWTTAF 964
Db       855 CLLPTLQVVSATEYPH----VLH--QQSITSVDDYLSSIKNKQALIVQTTLKEDIATIG- 907
Qy       965 TSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGN---WSGLNGGKNVCPLFT-FD 1020
Db       908 ----------KHRPIVTVPLVVNKYTGINGNTQIFQCGNLGYFMGRGVDRNLIPDSSGFR 957
Qy      1021 RTRRFIIACPRGGFICPVTG-----PSSGNRET----------TLSDQVRGIIVSGGAM 1064
Db       958 RQNNSAYMRKRHIFMTPIVANLLKRTSNLNNLTFEVETIRKNVQTIFDDKDNLNIFDNVV 1017

Qy      1065 VQLAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALE 1124
Db      1018 IEL------VKGLGESCENMTEDDLQFYLGDYYIMSDEIWSRFQLLTDSGTPWSVE---S 1068
Qy      1125 AVKILDEKTTAGDGETPTNLAF 1146
Db      1069 VTKILGSRK-----QEDCNLEF 1085
```

RESULT 12
S55600
single-stranded DNA binding protein 06 - equine herpesvirus 2
C;Species: equine herpesvirus 2
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 26-Aug-1999
C;Accession: S55600
R;Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A;Title: The DNA sequence of equine herpesvirus 2.
A;Reference number: S55594; MUID:95302501; PMID:7783207
A;Accession: S55600
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1145 <TEL>
A;Cross-references: GB:U20824; NID:g695172; PIDN:AAC13793.1; PID:g695178
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C;Superfamily: herpesvirus DNA-binding protein

Query Match 9.8%; Score 618; DB 2; Length 1145;
Best Local Similarity 23.1%; Pred. No. 3.8e-38;
Matches 295; Conservative 182; Mismatches 540; Indels 260; Gaps 47;

```
Qy         4 TQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLA 63
Db        17 TQASV----GPCGFVYLYPGDTFPVEEASLLGNLHAGGEVFSLPLLSGLTVEADFHVNVK 72
Qy        64 VVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFS 123
Db        73 AVHKKLDPA------TVSVKASAYHREVIVFANAACFKPIPAGPGLEGLCAASRQLFGYA 126
Qy       124 RCQGPPVDGAVETTGA-----EICTRLGLEP--ENTILYLVVTALFKEAVFMCNVFLHYG 176
Db       127 EFE--------ERAGGAARPFELADLGHLLPGAESHIAGVVVTESFKERLYRGQLVVVES 178
Qy       177 GLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHL 236
Db       179 QIQSVRVGECDAFKVPLYDGELFAKSPCR----------ENLRY---FYHAGVSRY 221
Qy       237 IHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARG 296
Db       222 LFEAHYTSLAQALRVRDVPGLIGALERQSF---HDQYKLP-----KVYECREFPATGHRG 273
Qy       297 ARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETP-TDIKEWPMFIGM 355
Db       274 A--------------GDCSLTIVDSVATELAVSYGLSFLEVPQEGTALLSYDKWPIFEGC 319
Qy       356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415
Db       320 ETPEQRVEALTQFNAKQAVHVHSQLFSGNSVLYLARVQKQASN--RGGGGENVYNSFFMG 377
Qy       416 AGPHLAANPQTDRDG-----HVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLER 470
Db       378 HGLACLAEPTQKENGLPSFPGVPASALSGSN---YSLHHLAYAASFSPQMLARHCYYLQF 434
Qy       471 CDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPR-FGQATRQ 529
Db       435 AQHQKSSNNSGYNVPTYVGTAANTPMCELCRGSCPASCVNTLFYRLRDRFPPVVASVRRD 494
Qy       530 PIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQER 589
Db       495 PYVVTGVAGA-YNDLDIAGNFANY---RDKDEESNQSEEREKFTYWQVTQTVL-----ER 545
Qy       590 LLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEA 649
```

```
Db        546 LSEAGICEGGEDVGDAIHNIGSFLKVFKEIDGIVDGEVARFINSMVK-NNVNYRESIKSI 604

Qy        650 THSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQ-------------- 695

Db        605 HHIVQYVCNVYWQPPCPVFLNLYYRCVLAVVQDICLPT--CMMYEQENPAVGVSPGEWLK 662

Qy        696 ---QVEGRNFRN-----------QFQPVLRRRFVDLFN------GGFISTRSITVTLSEG 735

Db        663 MHYQTLWTNFKNSCIDKGVLTGTEYKVVHKDQFCDFFDVDSAARGEFVSCKT-QVRISRA 721

Qy        736 PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVGL 795

Db        722 LMMAP-------------------------RVMKIKNRIIFSNSPGTES---------I 746

Qy        796 ASAYQRQEKRVD--MLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFW-TLLQRNQMPA 852

Db        747 QNAFVRGTPKGDSCVVSGPYMRFLSTYHSQLFP------GAKISPLFLWHTFSKKRQLPV 800

Qy        853 -DKLTHEEITTIA-AVKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQY--- 907

Db        801 FPNVPRESVTELANYVEQNSRLHGETSIIDVVPENFYTYAKVRLNNALFRACGQTQFYAT 860

Qy        908 LINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFTST 967

Db        861 TIHCLTPKIQTVPAEEYPHAL------GARGVADVGEYLGAARELT--VP---TVQCTSR 909

Qy        968 HLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGN---WSGLNGGKNVCPLFTFDRTRR 1024

Db        910 DNICEVGKCRPIVTLPLVVNKYTGVTGNSQIFQCANLGYFIGRGVDKNLIP--------- 960

Qy       1025 FIIACPRGGFICPVTGPSSGNRE-----TTLSDQVRGIIVSGGAMVQLAIYATVVRAVGA 1079

Db        961 -----DAGSF--KKQGVSTSMRKRHVFMTPLSDHLLRRSVQGAAV------AFEIEGVRR 1007

Qy       1080 RAQ---------HMAFDDWLSLT------------DDEF-------LARDLEELHDQII 1110

Db       1008 RVQQILSDGDNPHVIRDVVLQLVKSLGSECRSVSEYDLEYYMGQYYIFAGDVSE-RLQRL 1066

Qy       1111 QTLETPWTVEGALEAVKILDEKTTAGDGETPTNLAFNFDSCE-------PSHDTTSNVLN 1163

Db       1067 SDLGGDWSEEWALSVL---------GEEEDPLGGELEFEKVEDAECLGHPQQDEFALAPQ 1117

Qy       1164 ISGSNISG-STVPGLKR 1179

Db       1118 AAAPQYSGSSSVAGKKR 1134
```

```
RESULT 13
ONBEM1
ONA-binding protein - saimiriine herpesvirus 1 (strain 11)
C;Species: saimiriine herpesvirus 1
A;Note: host Saimiri sciureus (common squirrel monkey)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C;Accession: G36806
R;Albrecht, J.
submitted to the EMBL Data Library, January 1992
A;Description: Primary structure of the herpesvirus saimiri genome.
A;Reference number: A36806
A;Accession: G36806
A;Molecule type: DNA
A;Residues: 1-1128 <ALB>
A;Cross-references: GB:X64346; NID:g60320; PIDN:CAA45629.1; PID:g60327
R;Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; Wi
J. Virol. 66, 5047-5058, 1992
A;Title: Primary structure of the herpesvirus saimiri genome.
A;Reference number: A37309; MUID:92333688; PMID:1321287
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 6
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding
```

```
Query Match             9.6%;  Score 604;  DB 1;  Length 1128;
Best Local Similarity   22.9%;  Pred. No. 4.3e-37;
Matches  275;  Conservative  189;  Mismatches  509;  Indels  226;  Gaps   40;

Qy         11 PTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGART 70

Db         18 PVEPCGYIYVYPKEGFPFKEASLLGNKNVGASAMSLPLLSDLTVESNFSFNVKAVHKKID 77

Qy         71 TGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPPV 130

Db         78 M------TTLLVRVSAYHREAIVFFNTDLFEPIFVGPGLDILCSDARSLFGYTNF--VPR 129

Qy        131 DGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCN-VFLHYGGLDIVHINHGDVI 189

Db        130 TDLRDTVDIKDLYAPFYSEDSCFMAVVVTEGFKERLYFGNLVPIIAQGLK-VQINGREAV 188

Qy        190 RIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVAL 249

Db        189 KIPLYDEDLFSKSHEHL------------------PRFYIPSVSKYLHDSVFTSIAQAL 229

Qy        250 RVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPS 309

Db        230 RIRDVESVIRASEKQSIQDQYKLAKI-------------------------VNSK---- 259

Qy        310 PSGGFERRLASIMAADTA-------LHAEVIFNTGI-YEETP------TDIKEWPMFIGM 355

Db        260 -----DFSLQSVKCQDASAFMVIDCIAAELAMSYGLSFLEAPQDPCAVLDYTSWPIFETA 314

Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415

Db        315 ETEEDRIKAIQDWNAMMSVHVYTHLFSTNSVLYLTKINKQTQSNKSEQN---VYNTYFMQ 371

Qy        416 AGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGA 475

Db        372 HGLAYAADATQRENGEPAFSGAPKFSGGTYTLYHLALASSFSPHLLARNCYYMQFCQHQK 431

Qy        476 FTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPR-FGQATRQPIGVF 534

Db        432 STTNANYSVPQYVGTAAASDLCELCQGTCPASCIHTLFYRLKDRFPPVLGSQRRDPYVVT 491

Qy        535 GTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRG 594

Db        492 G-VSGQYNDLDMLGNFATF---REKEDEAVQNAESEKYTYWQLIQNVV-----EKLSTMG 542

Qy        595 APCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMA 654

Db        543 VTEGTVG-SELITDIQSFLKTFRDIDNVVDSEVVKFMNCLVK-NNINFRETIKTVHHVLH 600

Qy        655 LTFDPYSGAFCPITNFLVKRTHLAVVQDLAL-----------------SQCHCVFYGQQV 697

Db        601 YCCNVFWQAPCAMFLNLFYKSVLAIIQDICLPIAMTYEQDNPSIGMMPSEWLKVHY--QT 658

Qy        698 EGRNFR-----------NQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQ 746

Db        659 IWTNFKSSCLDRGVLTGSEHKIVHTDMFCDFLN--------IDSALS----------GQ 699

Qy        747 DAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRV 806

Db        700 IVP--MKMQVRLAKALLTVPKTIKIKNRIVFSN--SSMTETIQSGFIKSAT-----KKDS 750

Qy        807 DMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQRNQMPA-DKLTHEEITTIA- 864

Db        751 YIVTGPYMKFLNSLHKVMF-----PNAKISALYLWHTFSQKKQLPVLPGISRENMVELAN 805

Qy        865 AVKRFTEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSI------ITG 918

Db        806 YVETSSKMHDDMNVLDIIPTTLLTYAKVRLNNTILRTCGQTQFYATTLQCLLPTLQTISA 865

Qy        919 ARRPR---DPS--SVLHWIR--KD-------VTSAADIETQAKALLEKTENLPELWTTAF 964

Db        866 TEYPHVLLDQSIMSVDHYLSSIKDKHALTVQTTLKEDIATVGK---------------- 908
```

```
Qy        965 TSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGN---WSGLNGGKNVCPLFT-FD 1020
Db        909 -----------QRPIVTVPLVVNKYTGINGNTQIFQCGNLGYFMGRGVDRNLIPDSTGFR 957

Qy       1021 RTRRFIIACPRGGFICP-----VTGPSSGNRET----TLSDQVRGIIVSGGAMVQLAIYA 1071
Db        958 RQNNSSYMRRRHVFMTPMVAHLVKKNSNLNNLTFEVETIRKNVQNIFEDKD---NLNIFD 1014

Qy       1072 TV----VRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAV 1126
Db       1015 NVVLELVKGLGDSCENITEDDLQFYLGEYYIMSDEIWSRFQIITDSGAPWSVENVTKVL 1073


RESULT 14
T03105
major single-stranded DNA binding protein - alcelaphine herpesvirus 1
C;Species: alcelaphine herpesvirus 1
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Nov-1999
C;Accession: T03105
R;Ensser, A.; Pflanz, R.; Fleckenstein, B.
J. Virol. 71, 6517-6525, 1997
A;Title: Primary structure of the alcelaphine herpesvirus 1 genome.
A;Reference number: Z14840; MUID:97404659; PMID:9261371
A;Accession: T03105
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1127 <ENS>
A;Cross-references: EMBL:AF005370; NID:g2337967; PIDN:AAC58057.1; PID:g2337973
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding

 Query Match             9.5%;  Score 596;  DB 2;  Length 1127;
 Best Local Similarity   22.8%;  Pred. No. 1.7e-36;
 Matches  281;  Conservative  209;  Mismatches  542;  Indels  202;  Gaps   43;

Qy         11 PTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGA-- 68
Db         20 PIGPCGFIYIYPETHFNFKEISLLGDKYAEAGAFSLPLLHGVTVEEAFVPNVKAVYKKID 79

Qy         69 RTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGP 128
Db         80 MTT------VSVKLST--FYNRAIIFHNVEKFESIFSGPGLGSLCKEACDLFGY--VPFT 129

Qy        129 PVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINHGDV 188
Db        130 PL-GEGSTDVSDICPPVWQE-KDAIMAVVITEGFKERLHLGKLIYLKSQMHSVMINKTEV 187

Qy        189 IRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVA 248
Db        188 YRIPLYDEDLF-------------TKKSSLRR-------LYLPAVSEYLYYTLYTSLAQS 227

Qy        249 LRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKP 308
Db        228 LRVHNAASLVEAIQEQFVHDKYKMAKLVSFKEYPLATVGACDTTL-------------- 272

Qy        309 SPSGGFERRLASIMAADTALHAEVIFNTGIYEETPT-DIKEWPMFIGMEGTLPRLNALGS 367
Db        273 --------MVIDAVAAELGLSYSLSFFEAPQEKTKVQDYYSWDIFASCETDSDRLEALSK 324

Qy        368 YTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPS-------FNRFYQFAGPHL 420
Db        325 WNALQAIHIHAQLFSTNSIYYVNRVARQ--------APIPNSKVEPNVYNSYYLQHGLAN 376

Qy        421 AANPQTDRDGHVLSSQSTGSS--NTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTG 478
Db        377 LCBETLFEDGSPAFTGAPASSLDGSSFTLQHLAYAAAFSPNLLARMCYYLQFCQHQKSTL 436

Qy        479 GHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTM 537
Db        437 NPAYNITEYVGSAANSPV-CSLCSGQCPCVCINTLFYRLKDRFPPVLQGSRRDPYVITGI 495

Qy        538 NSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPC 597
```

```
Db        496 TNVFNELDFLGNFASFRDKDEDQNQTE---ETPRYTYWQLNQTLTEKLEAAGLVD--SPV 550

Qy        598 SSEGL-SSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALT 656
Db        551 ADEGAGGSGSMNLEKFVRTFSDIDSLVDAEAAKFINTMIK-NNVNFKESIKGVSHVIQYN 609

Qy        657 FDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQP--VLRRRF 714
Db        610 CNTYWQAPCSLMLNLYYRSILTIIQDIALP------ISTVYESENPAQGYKPNEWLKLHY 663

Qy        715 VDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGD--------------LA 759
Db        664 QTLW------TNFKSFFIDKGVITGTEMKVVHAEQFSDFFDVDAATNNMYSPVKVQVRLA 717

Qy        760 RVSVEVIRDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVD--MLHGALGFLL 817
Db        718 RAQVLALKNIKVKNRILFSG--TSMSEHYQ-------NAFLKTANRRDNYILAGPYVKFL 768

Qy        818 KQFHGLLFPRGMPPNSKSPNPQWFWT-LLQRNQMPADKLTHEEITTIAAVKRF------- 869
Db        769 NSFHRQLF-----PNLKI-SCLYLWSNFCKKKQIPCVPGVSAE-----ALNKFFSYINNN 817

Qy        870 TEEYAAINFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVL 929
Db        818 SKQFEEVNMLDVVPDSYVTYAKQRLNNAILRACGQTQFYAVTIHSIFPKVQE----TCAL 873

Qy        930 HWIRKDVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAMN---QRPMVVLGISI 986
Db        874 EYPHVLGTSSVD---SVEDYVNNVQNLKAL----TVNSSLRESAANLARSRPIVTLPVVV 926

Qy        987 SKYHGAAGNNRVFQAGN---WSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTG--- 1040
Db        927 NKYTGIAGNAQLFQSANLGYFMGRGVDKNLLGDSLFVKKQQNSYMRKKYLFMTPLVGNLL 986

Qy       1041 -PSSGNRETT------------LSDQVRGIIVSGGAMVQLAIYATVVRAVGARAQHMAF 1086
Db        987 KPSYTHQGTAFEIETVKRTIQSILEDQADEDVLN-------RVVCELVKSLGAGCADLTL 1039

Qy       1087 DD-WLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGETPTNLA 1145
Db       1040 DDIQFYLGSYGMFSENILEKLDQ-LRELVGPWTHEWAESVLK-------SGTCETDEVQF 1091

Qy       1146 FNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKR 1179
Db       1092 VAFE---------EEQVKLTSMDHSGKVVGGKKR 1116


RESULT 15
T44001
major DNA binding protein [imported] - human herpesvirus 6
C;Species: human herpesvirus 6
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T44001; T44188
R;Isegawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; Kawar
J. Virol. 73, 8053-8063, 1999
A;Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B
A;Reference number: Z22732; MUID:99412319; PMID:10482554
A;Accession: T44001
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1132 <ISE>
A;Cross-references: EMBL:AB021506; NID:g4995977; PIDN:BAA78262.1; PID:g4996029
A;Experimental source: strain HST; pop. variant B
R;Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A;Title: Human herpesvirus 6B genome sequence: coding content and comparison with human H
A;Reference number: Z22734; MUID:99412318; PMID:10482553
A;Accession: T44188
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1132 <DOM>
A;Cross-references: EMBL:AF157706; PIDN:AAB06339.1
A;Experimental source: strain Z29; variant B
```

C;Genetics:
A;Gene: U41
C;Superfamily: herpesvirus DNA-binding protein

Query Match 7.8%; Score 490.5; DB 2; Length 1132;
Best Local Similarity 22.0%; Pred. No. 1.9e-28;
Matches 278; Conservative 180; Mismatches 535; Indels 273; Gaps 54;

```
Qy    2 ENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
Db    3 DENETVVSAPVSTAAWIYVFPKEKELLDVLSVLSLMERNSPVVISPLLMNLTVENDFSTT 62

Qy   62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVL-PSSAAPNLTRACNAARERF 120
Db   63 V-------KTPITNFGGTILTKITSFMPVCFFFHGTEQLVGMAEDHGDLIRLCEQTRQKF 115

Qy  121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
Db  116 HLQSFEVPTARKVIDIKA--LCSAVGKDADSVICHVACGNGFKELLF--------AGLLI 165

Qy  181 --------VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTG 232
Db  166 PCVEEQIQVQVGEYSCVKIPLYSATLFETE---------ETISLSSCTEFIQERGFFLPA 216

Qy  233 LCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTT 292
Db  217 LSETLFYYVFTSWGTTLRFSNTKELIDAGLKQFTQDGEQTVKLAPHKTYL---------- 266

Qy  293 TARGARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFN------TGIYEETPTDI 346
Db  267 ---GISGQKISAVEKD----------FLMLVDSVV-TELSFSHVAEYLDSVYD--PSQI 309

Qy  347 ---KEWPMFIGMEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVE---------D 394
Db  310 MNFNDWPIIRNSETHAERMAQLTNLKLHLSSHLAVLIFAPNSILYCSKLAFIPNVKQAFN 369

Qy  395 SGMTEAKDGGPGPSFNRFYQFAGPHLAANPQTDRDGH-VLSSQSTGSSNTEFSVDYLALI 453
Db  370 SVMTQ-------ELLLRSLSFCNALSSLTEDVYNDNRKIIKCDSTSGKDDKFSANHLAYA 422

Qy  454 CGFGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEI----------PCSLCEKH 503
Db  423 CATCPQLLSYVVWNLNRMSV-----------YNAGNAYTEIYNHLVNCSANLCEFCDGK 470

Qy  504 TRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQT 563
Db  471 CCQSCIGTAMVRVGTRLPAIPKNVKKEPLVMSMFSRYYAEVDILGSFG-----RKPVSEL 525

Qy  564 EAAKATMQDTYRATLERLFIDLEQERLLDRGA-------PCSSEGLSSVIVDHPTFR--- 613
Db  526 KEIGKDQQNTLS--------------LDRGKFVSQIFDYCKKNSLIDPVTGEDTFNVRS 570

Qy  614 -----RILDTLRARIEQTTTQFMKVLVETRDYKI-REGLSEATHSMALTFDPYSGAFCPI 667
Db  571 KKDFVSIIHGLTQCIEECVS---RCIVEMRRTQTPREQIENCLQSFNVDTTPYATAFSPF 627

Qy  668 TNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLF--------N 719
Db  628 LTFAYYKVILTVLQNLAL----IVASGHVVDRPCTGNSISKWLVQQYQSLYGTFHSSYLK 683

Qy  720 GGFISTRSITVTLSEGPVSAPNPTLGQ--DA---PAGR----TFDGDLARVSVEVIRDIR 770
Db  684 KGFLNTRTVKV--------ASNVDMEQILDCDLYKSGKYVKTTIQAKLCRLSMQCLRDFR 735

Qy  771 VKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMP 830
Db  736 IKNRPFNKSSKTAHNNPYFKKNV---------KHKKNPLSGCISFLLFKYHDKLF----- 781

Qy  831 PNSK-SPNPQWFWTLLQRNQMPADKLTHEEITT-IAAVKRFTEEYAAINFINLPPTCIGE 888
Db  782 PNVKISCLELWQRFLLNNVPKTLDIGNPEEVKTFIKFAFSITNTYDEIDIIDIQPECLST 841

Qy  889 LAQFYMANLILKYCDHSQYL--INTLTS-IITG---------ARRPRDPS--SVLHWIRK 934

                                                    ::|:  |    | :::|
Db  842 FIDCYFHNKFLSALGFHDYLTSLHGLTSKLVTQNPVLFPVVLDKQPKFSSIQEYLVYVKK 901

Qy  935 DVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAG 994
Db  902 LVLDG--VPNPVIASLSKEPN----FGTIFTS----------RSLVTFGLTLEKFVSLA- 944

Qy  995 NNRVFQAG--NWSGLNG-GKNVCPL------FTFDRTRRFIIACPRGGFICPVTGPSSGN 1045
Db  945 NREYFQFGQLGWIGGSGVDRNLNPTSSALQDFRFMR-QKTIIATKFSEVIV-----KKVR 998

Qy 1046 RETTLSDQ--VRGIIVS------GGAMVQLAIYATVVRAVGARAQHMAFDDWLSLTDD-E 1096
Db  999 RETIMFDTEVVKGKVLSIVENLTNDIDPELLIIAEVMR---DREDKPTMDDMLFFVDGRE 1055

Qy 1097 FLARD----LEELHDQIIQTLETPWTVEGALEAV----------KILDEKTTAGDG--- 1138
Db 1056 ALAASIMLKLNHLVDMNVRDFSIA-NLQSVFEAVSSNDAPVYDFSEILAEEDDQASGVLK 1114

Qy 1139 --ETPT 1142
Db 1115 CDETET 1120
```

Search completed: January 30, 2004, 13:14:42
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:06:32 ; Search time 47 Seconds
(without alignments)
4062.725 Million cell updates/sec

Title: US-09-769-699-2
Perfect score: 6294
Sequence: 1 MENTQKTVTVPTGPLGYVYA.....DELFDLSGIPIKHGNITMEM 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6294	100.0	1203	22	AAM52633	Varicella-Zoster v
2	3054	48.5	1452	23	AAE25598	Herpesvirus ICP8-G
3	3054	48.5	1452	24	ABG73090	Herpes simplex vir
4	3016.5	47.9	1196	19	AAW72105	HSV-2 strain SB5 C
5	1566.5	24.9	623	19	AAW72197	HSV-2 strain SB5 C
6	1380.5	21.9	543	19	AAW72196	HSV-2 strain SB5 C
7	620	9.9	1132	21	AAB53126	Macaca mulatta rha
8	183	2.9	35	20	AAY26935	Varicella zoster v
9	183	2.9	35	22	AAB73931	Varicella zoster V
10	156.5	2.5	1841	21	AAB07562	Protein encoded by
11	125.5	2.0	19938	24	ABP76682	Streptomyces virid
12	123.5	2.0	3210	22	AAB73958	Mycelia sterilia c
13	123	2.0	1074	20	AAY00206	Enterococcus faeca
14	123	2.0	1074	20	AAY00188	Enterococcus faeca
15	123	2.0	1074	23	ABP43407	E faecalis EF094 p
16	123	2.0	1074	23	ABP43425	E faecalis EF102 p
17	123	2.0	1074	24	ABU13686	Enterococcus faeca
18	123	2.0	1074	24	ABU13704	Enterococcus faeca
19	122.5	1.9	901	22	ABB58224	Drosophila melanog
20	122.5	1.9	2387	21	AAY53665	Mechanical stress
21	121	1.9	1069	21	AAB41936	Human ORFX ORF1700
22	121	1.9	3158	23	ABP62763	S. roseosporus dap
23	120.5	1.9	2597	21	AAY53664	Mechanical stress
24	120.5	1.9	2597	23	ABG32889	Rat osteoclast pro
25	120.5	1.9	2597	23	ABG32890	Rat osteoclast pro
26	120.5	1.9	2597	23	ABG31316	Rat protein 608.
27	120.5	1.9	2597	23	AAB47933	Rat OCP. Rattus r
28	118.5	1.9	1076	22	ABG20205	Novel human diagno
29	118	1.9	1784	22	ABG24772	Novel human diagno
30	115	1.8	1150	14	AAR43950	Ge protein fragmen
31	114.5	1.8	1496	20	AAW81030	Melanoma associate
32	114.5	1.8	1496	21	AAY70469	Human p53 target m
33	114.5	1.8	1496	24	ABU03498	Angiogenesis-assoc
34	114.5	1.8	1498	22	ABB11587	Human peroxidasin
35	114	1.8	774	7	AAP60616	Cephalosporin C ac
36	111	1.8	4630	18	AAW19629	Streptomyces venez
37	111	1.8	4630	21	AAY77177	S. venezuelae vep
38	110.5	1.8	1532	22	ABB68585	Drosophila melanog
39	110.5	1.8	3070	22	AAG81195	Mycobacterium tube
40	109.5	1.7	2630	24	ABG76186	Human serine/threo
41	109.5	1.7	7968	24	ABG76187	Human serine/threo
42	109	1.7	11096	22	AAE10129	Streptomyces nours
43	108.5	1.7	1451	22	AAB95868	Human protein sequ
44	108	1.7	2596	22	AAB30569	A splice variant o
45	106.5	1.7	774	16	AAR86374	Cephalosporin C ac

ALIGNMENTS

RESULT 1
AAM52633
ID AAM52633 standard; Protein; 1203 AA.
XX
AC AAM52633;
XX
DT 18-FEB-2002 (first entry)
XX
DE Varicella-Zoster virus (VZV) ORF29p protein.
XX
KW ORF29p; open reading frame 29p protein; Varicella-Zoster virus; VZV;
KW drug delivery; gene delivery; fusion protein; protein secretion.
XX
OS Human herpesvirus 3.
XX
PN WO200154709-A1.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02500.
XX
PR 25-JAN-2000; 2000US-0177901.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Silverstein S, Annunziato P, Gershon A, Lungu O;
XX
DR WPI; 2001-488744/53.
DR N-PSDB; ABA02175.
XX
PT Composition of matter for delivering desired compounds into eukaryotic
PT cell, comprises Varicella-Zoster virus open reading frame 29p protein

09/769,699
2-27-04
Search Notes

PT either bound to an agent or operably affixed to lipid-soluble group -
XX
PS Claim 1; Fig 6; 68pp; English.
XX
CC This sequence represents the Varicella-Zoster virus (VZV) open
CC reading frame 29p (ORF29p) protein. Although ORF29p is the major DNA
CC binding protein of VZV, the invention is based on the discovery that the
CC VZV ORF29p protein can readily enter and exit eukaryotic cells.
CC Accordingly, the invention relates to a composition for delivery of an
CC agent into a eukaryotic cell, comprising the VZV ORF29p protein bound to
CC the agent to be delivered. The agent delivered may be a polypeptide, a
CC polynucleotide, or an organic compound. The invention also relates a
CC composition in which the ORF29p protein is linked to a lipid soluble
CC group that permits the protein to be anchored to a lipid membrane, and
CC lipid vesicles containing this composition. The invention additionally
CC relates to the use of the ORF29p protein as a facilitator of secretion of
CC a desired protein, in which the desired protein and the ORF29p protein
CC are recombinantly expressed as a fusion protein. The invention further
CC encompasses nucleic acids encoding the VZV ORF29p protein and their use
CC in detection of VZV ORF29p nucleic acids, and a monoclonal antibody
CC against the ORF29p protein. Compositions of the invention are useful for
CC the delivery of a prophylactic or therapeutic agent to a eukaryotic cell,
CC particularly a human cell. Examples of agents that can be delivered to a
CC cell include proteins such as insulin, factor VIII, factor IX, and
CC proteases; polynucleotides (e.g., for use in gene therapy), and organic
CC compounds such as vitamins and a wide variety of pharmaceuticals (e.g.,
CC antineoplastic, antidiabetic or immunosuppressive drugs).
XX
SQ Sequence 1203 AA;

Query Match 100.0%; Score 6294; DB 22; Length 1203;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60

Qy         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120

Qy        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDI 180

Qy        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDC 240

Qy        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRN 300

Qy        301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLP 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLP 360

Qy        361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 RLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHL 420

Qy        421 AANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTGGH 480

Qy        481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNSQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVFGTMNSQ 540

Qy        541 YSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRGAPCSSE 600

Qy        601 GLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660

Qy        661 SGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFNG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRRFVDLFNG 720

Qy        721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGN 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGN 780

Qy        781 CTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQW 840

Qy        841 FWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 FWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900

Qy        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960

Qy        961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFD 1020

Qy       1021 RTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGAR 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 RTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGAR 1080

Qy       1081 AQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGET 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGET 1140

Qy       1141 PTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHGNIT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 PTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFDLSGIPIKHGNIT 1200

Qy       1201 MEM 1203
              |||
Db       1201 MEM 1203
```

RESULT 2
AAE25598
ID AAE25598 standard; Protein; 1452 AA.
XX
AC AAE25598;
XX
DT 04-NOV-2002 (first entry)
XX
DE Herpesvirus ICP8-GFP fusion protein.
XX
KW ICP8 protein; herpesvirus; green fluorescent protein; GFP; infection;
KW antiviral agent; fusion protein.
XX
OS Chimeric - Herpesvirus.
OS Chimeric - Unidentified.
XX
PN US6399354-B1.
XX
PD 04-JUN-2002.
XX
PF 31-JUL-1998; 98US-0127227.
XX
PR 31-JUL-1998; 98US-0127227.
XX
PA (HARD) HARVARD COLLEGE.
XX
PI Knipe DM, Taylor TJ, Mcnamee EE;

CC
DR WPI; 2002-546283/58.
DR N-PSDB; AAD41974.
CC
PT Novel fusion protein comprising an ICP8 herpes viral protein and a
PT detectable protein, useful for screening virus resistant cells and for
PT identifying antiviral agents or agents that block the viral protein
PT expression -
CC
PS Claim 9; Column 21-28; 29pp; English.
CC
CC The invention relates to a fusion protein which comprises an ICP8
CC herpesviral protein and a detectable protein e.g. green fluorescent
CC protein (GFP). Fusion proteins of the invention are useful for
CC screening virus resistant cells and for identifying antiviral agents
CC or agents that block the viral protein expression, for assaying the
CC efficiency of antiviral agents or immunological reagents, to screen
CC for infection in cells, to detect virus-resistant and/or virus-
CC susceptible cells, and to identify antiviral drug targets, antiviral
CC agents or compounds, and/or agents that block the viral protein
CC expression. The present sequence is ICP8-GFP fusion protein.
CC
SQ Sequence 1452 AA;

Query Match 48.5%; Score 3054; DB 23; Length 1452;
Best Local Similarity 49.8%; Pred. No. 2.5e-288;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

```
Qy          1 MENTQKTVT---VPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVE 55
              ||   || |    || |||||||  ||  | ::|  :: |:||| |:|:|: ||:  ||||
Db          1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSGDADVAVAPLVVGLTVE 58

Qy         56 KTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNA 115
                | :::||| |:|||||| |  ::|||| ||:  ||:|||||:|: ||: |||||| |
Db         59 SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCER 118

Qy        116 ARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175
              ||  ||||      | |   |||| :| ||||:|:  :|||||  ||||| : | |||
Db        119 ARRHFGFSDYTPRPGDLKHETTGEALCERLGLDDPDRALLYLVVTEGFKEAVCINNTFLHL 178

Qy        176 GGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCH 235
              || | | |   :| |||::|:|||||| :|:: :||| :|||||| | || ||:|  |
Db        179 GGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFNRPLNR 238

Qy        236 LIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTAR 295
              |: : |: | ||||| ||| |||| |||||||||||||| || |||:| |::|     | |
Db        239 LLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KTPR 296

Qy        296 GARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355
              | |           |:||||:||||:|| | ||  | | :  :::| |||| ||:  |
Db        297 GGR-----DGGGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLCEGQ 351

Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415
              :    | ||:|:| || ||::|||||| ||||:||||:|:|  : ||    ||| ||:
Db        352 DTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLV 410

Qy        416 AGPHLAANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYL 468
              | |:|||| ||:|||:       ::   | :  ||: ::||::|||   |||::||||
Db        411 PGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYL 469

Qy        469 ERCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQAT 527
              |||| |  |    | :||  :  :::||:||    || |||| ||| | |:|  |
Db        470 ERCDGGVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAA 529

Qy        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
              |  |||||||| ||||| |||||  :  |::  | :| |:  ||:|||  ||:  :||
Db        530 RGAIGVFGTMNSMYSDCDVLGNYAAFSALKR-ADGSETARTIMQETYRAATERVMAELET 588

Qy        588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647
              : :|:  | :   | ::| :      ::: :|  :::   | |: ||| |:| |:|
Db        589 LQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLG 648

Qy        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707
```

```
Db        649 EANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQ 708

Qy        708 PVLRRRFVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766
              |||||| :|:|| ||:| ::|| |||  : || : ||| |||  :|:|||:|| ::
Db        709 PVLRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFP 768

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826
              :::|||:|||:|  | ||||:||:  | ||||: :||||:|  | |||||||||  :||
Db        769 KELRVKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFP 828

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886
              | || |  |||||||| |||||:||  |: |:| ||| :|:|: :| |||||| |  :
Db        829 NGKPPGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFINLAPNNV 888

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
              ||| :|||| ||:|||||  | ||||||:|| |:|||     :  |          |:|  |
Db        889 SELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAW---SAQGGAGLEAGA 945

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
              :||::  :  |  ||: | | :|:|  | ||||||:|||||: |||:|||||||:
Db        946 RALMDAVDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWAS 1005

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
              | |||| ||| ||||||:::|||| ||:  :  |  |::| :|:|||  |||  |
Db       1006 LMGGKNACPLLIFDRTRKFVLACPRAGFVCAASNLGGGAHESSLCEQLRGIISEGGAAVA 1065

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL--- 1123
                :::   |:::| | | :  :|||: :||:|: :: ||  : ::     |: : ||
Db       1066 SSVFVATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVA 1125

Qy       1124 -EAVKILDEKTTAGDGETPTNLAFNFD--SCEPSHDTTSNVLNISGSNISGSTVPGLKRP 1180
              ||  :: :  ||:          |||  ||      |     |  |  | | ||
Db       1126 HEAEALVSQLGNAGE-------VFNFGDFGCE-----DDNATPFGGPGAPGPAFAGRKRA 1173

Qy       1181 PEDDELFDLSGIPIKHGNITMEM 1203
                 |:|   | || | |::|:|
Db       1174 FHGDDPFG-EGPPDKKGDLTLDM 1195


RESULT 3
ABG73090
ID   ABG73090 standard; Protein; 1452 AA.
XX
AC   ABG73090;
XX
DT   14-APR-2003  (first entry)
XX
DE   Herpes simplex virus (HSV) ICP8-GFP fusion protein.
XX
KW   Herpesvirus; ICP8-GFP; green fluorescent protein; virucide;
KW   fusion protein; HSV.
XX
OS   Herpes simplex virus.
OS   Synthetic.
XX
PN   US2002151033-A1.
XX
PD   17-OCT-2002.
XX
PF   16-JAN-2002; 2002US-0050673.
XX
PR   31-JUL-1998;   98US-0127227.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Knipe DM,  Taylor TJ,  McNamee EE;
XX
DR   WPI; 2003-182512/18.
DR   N-PSDB; ABX15344.
XX
PT   Novel fusion protein useful for detecting the presence, absence or
PT   amount of viral DNA replication and infection and as marker, comprises
```

PT viral protein derived from virus and detectable protein -
XX
PS Disclosure; Fig 7; 30pp; English.
XX
CC The invention relates to a fusion protein comprising a viral protein with
CC at least one function and a detectable protein, fused to maintain the
CC open reading frame. The fusion protein is useful for determining whether
CC a cell is a virus-resistance cell or a virus-susceptible cell, by
CC contacting a virus that expresses the fusion protein with the cell to be
CC tested, under conditions sufficient for the virus to infect the cell, and
CC detecting the presence or absence of the fusion protein, where the
CC absence of the fusion protein identifies a virus-resistant cell and the
CC presence of the fusion protein identifies a virus-susceptible cell. The
CC fusion protein is useful for identifying an anti-viral agent or an agent
CC that blocks the expression of the fusion protein, by contacting a virus
CC that expresses the protein, a host cell and the agent to be tested, in
CC conditions sufficient to allow for the virus to infect the cell, and
CC detecting the amount of the virus present, where a decrease in the amount
CC of virus present identifies the agent. This sequence represents an
CC ICP8-green fluorescent protein (ICP8-GFP) fusion protein of the
CC invention.
XX
SQ Sequence 1452 AA;

Query Match 48.5%; Score 3054; DB 24; Length 1452;
Best Local Similarity 49.8%; Pred. No. 2.5e-288;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

```
Qy          1 MENTQKTVT---VPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVE 55
              ||   || |   || |||||||  ||  | ::|  :: |:||| |:|:|: ||:  ||||
Db          1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSGDADVAVAPLVVGLTVE 58

Qy         56 KTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNA 115
                | :::||| |:|||||  |  ::|||| ||:  ||:||||:|: ||: |||||| |
Db         59 SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCER 118

Qy        116 ARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175
              ||  ||||     | |   |||| :| ||||:|:  :|||||  ||||| : | |||
Db        119 ARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINNTFLHL 178

Qy        176 GGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCH 235
              || | | |    :| |||::|:|||||| :|:: :||| :|||||| | || |||:|
Db        179 GGSDKVTIGGAEVHRIPVYPLQLFMPDFSRVIAEPFNANHRSIGENFTYPLPFFNRPLNR 238

Qy        236 LIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTAR 295
              |: : |: | ||||| ||| |||| ||||||||||||| || |||:| |::     | |
Db        239 LLFEAVVGPAAVALRCRNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KTPR 296

Qy        296 GARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355
              | |          |:|||||:|||:||  | ||  | |: :::| |||| ||:  |
Db        297 GGR-----DGGGKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPTDISAWPLCEGQ 351

Qy        356 EGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQF 415
              :    | ||:|:|  || ||:|||||| ||||:|||:|:|  : ||    ||| ||:
Db        352 DTAAARANAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLV 410

Qy        416 AGPHLAANPQTDRDGHVL------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYL 468
              | |:||||| ||:|||:        ::   | :  ||: ::||::|||   |||::||||
Db        411 PGTHVAANPQVDREGHVVPGFEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYL 469

Qy        469 ERCDAGAFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQAT 527
              |||| |   |   :||  :  :::||:||   ||  | ||||: ||| | |:  |
Db        470 ERCDGGVIVGRQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTTLMRLRARHPKFASAA 529

Qy        528 RQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQ 587
              |  |||||||||| ||||| |||||  :  |::  | :| |:  ||:|||| ||:  :||
Db        530 RGAIGVFGTMNSMYSDCDVLGNYAAFSALKR-ADGSETARTIMQETYRAATERVMAELET 588

Qy        588 ERLLDRGAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLS 647
              : :|:  | :   | ::| :      ::: :|  :::   | |: ||| |::| |:||
Db        589 LQYVDQAVPTAMGRLETIITNREALHTVVNNVRQVVDREVEQLMRNLVEGRNFKFRDGLG 648

Qy        648 EATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQ 707
              || |:|:|| |||:   ||:  | :|:::||| |||||||| || || ||||||||||
Db        649 EANHAMSLTLDPYACGPCPLLQLLGRRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQ 708

Qy        708 PVLRRRFVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766
              |||||| :|:|| ||:| :::|| |||| : ||:| | || |||  :|:|||||::
Db        709 PVLRRRVMDMFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFP 768

Qy        767 RDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFP 826
              ::|||:||:|:|   | ||||:||:  | |||| :|||:| | |||||||| :||
Db        769 KELRVKSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAAIFP 828

Qy        827 RGMPPNSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCI 886
              | || | |||||||| |||||:||  |: |:| ||| :|:| : |||||||  |  :
Db        829 NGKPPGSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKKFSLDYGAINFINLAPNNV 888

Qy        887 GELAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQA 946
              ||| :|||| ||:|||| | ||||:|| |:|||     :   |         | :   |
Db        889 SELAMYYMANQILRYCDHSTYFINTLTAIIAGSRRPPSVQAAAAW---SAQGGAGLEAGA 945

Qy        947 KALLEKTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSG 1006
              :||::  :  |  ||:  | :|:|  |  |||||||:|||||:| |||:|||||||:
Db        946 RALMDAVDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWAS 1005

Qy       1007 LNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQ 1066
              | |||| ||| |||||:|:|||| ||:|  :     |  |:: | :|:|||   ||| |
Db       1006 LMGGKNACPLLIFDRTRKFVLACPRAGFVCAASNLGGGAHESSLCEQLRGIISEGGAAVA 1065

Qy       1067 LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL--- 1123
                 :::   |:::| | | :  :|||: |||:|: :: ||  : ::     |: : ||
Db       1066 SSVFVATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMELTARALERGNGEWSTDAALEVA 1125

Qy       1124 -EAVKILDEKTTAGDGETPTNLAFNFD--SCEPSHDTTSNVLNISGSNISGSTVPGLKRP 1180
              || :: :   ||:          |||   ||  |  |  |  |  |  |  | | |||
Db       1126 HEAEALVSQLGNAGE-------VFNFGDFGCE-----DDNATPFGGPGAPGPAFAGRKRA 1173

Qy       1181 PEDDELFDLSGIPIKHGNITMEM 1203
                 |: |  | | |  |::|::|
Db       1174 FHGDDPFG-EGPPDKKGDLTLDM 1195
```

RESULT 4
AAW72105
ID AAW72105 standard; Protein; 1196 AA.
XX
AC AAW72105;
XX
DT 18-DEC-1998 (first entry)
XX
DE HSV-2 strain SB5 Contig ID 12 ORF#2 protein.
XX
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
XX
OS Herpes simplex virus type 2.
XX
PN WO9820016-A1.
XX
PD 14-MAY-1998.
XX
PF 31-OCT-1997; 97WO-US20016.
XX
PR 09-JUN-1997; 97US-0049018.
PR 04-NOV-1996; 96US-0030279.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
PI Esser KM, Leary JJ;
XX
DR WPI; 1998-286847/25.
DR N-PSDB; AAV62156.
XX
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention

```
PT   and treatment of infection or inducing immunological response in
PT   mammal
XX
PS   Claim 10; Page 83-84; 748pp; English.
XX
CC   This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC   sequence of the invention. This sequence was isolated from a HSV-2 strain
CC   SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 12.
CC   The proteins can be used for the treatment or prevention of disease, to
CC   induce an immunological response in a mammal or to identify inhibitors,
CC   activators or novel antivirals. Antagonists of the proteins can be used
CC   to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC   it can also be used to induce an immunological response in a mammal.
XX
SQ   Sequence   1196 AA;

 Query Match             47.9%;  Score 3016.5;  DB 19;  Length 1196;
 Best Local Similarity   49.1%;  Pred. No. 8.4e-285;
 Matches  596;  Conservative  211;  Mismatches  367;  Indels   41;  Gaps   16;

Qy          4 TQKTVTVPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
              |   ||  ||  ||:||||   ||   |  |:|  :| |:||| |:|:|: ||:  |||| | ::
Db          7 TTTTVKVPPGPMGYVYGRACPAEGLEL--LSLLSARSGDADVAVAPLIVGLTVESGFEAN 64

Qy         62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFG 121
              :| | |:|||||  |   ::|||   ||: |||:|||||:|: ||: ||||||  |   ||  ||
Db         65 VAAVVGSRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQAPNLTRLCERARRHFG 124

Qy        122 FSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIV 181
              ||      | |   ||||  :| ||||:|:  :||||:|  |:||| : | ||| ||:| |
Db        125 FSDYAPRPCDLKHETTGDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGGMDKV 184

Qy        182 HINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCV 241
              |   :| |||::|:|:|||| :|:: |||| :|||||  | || ||:|  |  |: : |
Db        185 TIGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAV 244

Qy        242 IAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRND 301
              : | ||||| ||| |||| |||||||||||||| || |||:| |::|       ||||
Db        245 VGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KPQRGAR--- 299

Qy        302 VNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPR 361
                    |:||||:||||:|| | ||  | | :  :::| | ||   ||:  | |     |
Db        300 --DAGNKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPPDITTWPLLEGQETPAAR 357

Qy        362 LNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLA 421
              |:|:| || ||::|||||| ||||:||||:|:|  : ||    ||| ||:   | |:|
Db        358 AGAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVA 416

Qy        422 ANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAG 474
              |||| ||:|||:       ::   | :  ||: ::||::|||   |||::||||||| |
Db        417 ANPQLDREGHVVPGYEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYLERCDGG 475

Qy        475 AFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGV 533
              |    |  :||  :  :::||:||   ||  |||||: ||| | |:|  | |  |||
Db        476 VIVGRQEMDVFRYVADSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGV 535

Qy        534 FGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDR 593
              |||||| ||||| ||||| :  |::  | :|  :  ||:|||  ||:  :||  : :|:
Db        536 FGTMNSAYSDCDVLGNYAAFSALKR-ADGSENTRTIMQETYRAATERVMAELEALQYVDQ 594

Qy        594 GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSM 653
              | :   | ::|         ::: ::  :::   | |: |:| |::| |:||:|| |:|
Db        595 AVPTALGRLETIIGTREALHTVVNNIKQLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654

Qy        654 ALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRR 713
              :|: |||:    ||:    | :|::||| |||||||| || || ||||||||||||||| |
Db        655 SLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRRR 714

Qy        714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
              :|||| ||:| :::|| ||||  : ||: | || |||  :|:||:|||::   :::|||
Db        715 VMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVK 774

Qy        773 NRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPN 832
```

```
              :||:|:|   | ||||:||:  | ||||: :||||:| | |||||||||| ::|| | ||
Db        775 SRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPPG 834

Qy        833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
              |  |||||||| |||||:||  |: |:| ||| :|||: :| |||||||  |  : ||| :
Db        835 SNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMY 894

Qy        893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
              |||| ||:|||||  | |||||::| |:|||   :  |  :    | :|  |:||::
Db        895 YMANQILRYCDHSTYFINTLTAVIAGSRRPPGVQAAAAWAPQ---GGAGLEAGARALMDS 951

Qy        953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
              :  |  ||: | | :|:|  |  |||||||:|||||:|||:||||||||: | ||||
Db        952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGKN 1011

Qy       1013 VCPLPTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYAT 1072
              |||  |||||:|::|||| ||:|  :   |  | :|:|||  ||| |  :::
Db       1012 ACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071

Qy       1073 VVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVKI 1128
              |:::| | | :  :|||:| :||:|: :: |  : ::     |: : ||    ||  :
Db       1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEAL 1131

Qy       1129 LDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFD 1188
              : :   ||:          |||  |  ::  :: :  :|:    ||    |: |
Db       1132 VSQLGAAGE-------VFNFGDFGDEDDHAASFGGLAAA--AGAAGVARKRAFHGDDPFG 1182

Qy       1189 LSGIPIKHGNITMEM 1203
                | | |  ::|::|
Db       1183 -EGPPEKK-DLTLDM 1195


RESULT 5
AAW72197
ID   AAW72197 standard; Protein; 623 AA.
XX
AC   AAW72197;
XX
DT   13-JAN-1999  (first entry)
XX
DE   HSV-2 strain SB5 Contig ID 15 ORF#32 protein.
XX
KW   HSV-2 strain SB5; immunological response induction; therapy;
KW   antiviral identification; viral protein inhibitor.
XX
OS   Herpes simplex virus type 2.
XX
PN   WO9820016-A1.
XX
PD   14-MAY-1998.
XX
PF   31-OCT-1997;   97WO-US20016.
XX
PR   09-JUN-1997;   97US-0049018.
PR   04-NOV-1996;   96US-0030279.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Chan JY,  Dabrowski-Amaral CE,  Delvecchio AM,  Dillon SB;
PI   Esser KM,  Leary JJ;
XX
DR   WPI; 1998-286847/25.
DR   N-PSDB; AAV62176.
XX
PT   Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT   and treatment of infection or inducing immunological response in
PT   mammal
XX
PS   Claim 10; Page 123; 748pp; English.
XX
CC   This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC   sequence of the invention. This sequence was isolated from a HSV-2 strain
CC   SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
```

```
CC   Based on homology, this sequence is a major DNA-binding protein.
CC   The proteins can be used for the treatment or prevention of disease, to
CC   induce an immunological response in a mammal or to identify inhibitors,
CC   activators or novel antivirals. Antagonists of the proteins can be used
CC   to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC   it can also be used to induce an immunological response in a mammal.
XX
SQ   Sequence   623 AA;

  Query Match             24.9%;  Score 1566.5;  DB 19;  Length 623;
  Best Local Similarity   50.0%;  Pred. No. 2.2e-143;
  Matches  313;  Conservative  100;  Mismatches  192;  Indels   21;  Gaps    9;

Qy          4 TQKTVTVPTGPLGYVY-ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSL 62
              |  || || ||:|||| ||  | |:|  :| |:||| |:|:|: ||:  |||| | :::
Db          7 TTTTVKVPPGPMGYVYRACPAEGLEL--LSLLSARSGDADVAVAPLIVGLTVESGFEANV 64

Qy         63 AVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGF 122
              | | |:|||| |  ::||  ||: |||:|||||:|: ||: |||||| |  ||  |||
Db         65 AAVVGSRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQAPNLTRLCERARRHFGF 124

Qy        123 SRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVH 182
              |     | |   |||| :| ||||:|:  :||||:|  |:||| : | ||| ||:| |
Db        125 SDYAPRPCDLXHETTGDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGGMDKVT 184

Qy        183 INHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVI 242
              |   :| |||::|:|:|||| :|:: |||| :|||||| | || ||:|  |  |: : |:
Db        185 IGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAVV 244

Qy        243 APMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDV 302
              | |||||| ||| |||| |||||||||||||| || |||:| |::|       ||||
Db        245 GPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KPQRGAR---- 298

Qy        303 NSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPRL 362
                |:||||:||||:|| | ||  | | :  :::| | ||  ||:  | |    |
Db        299 -DAGNKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPPDITTWPLLEGQETPAARA 357

Qy        363 NALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLAA 422
              |:|:| || ||::|||||| ||||:||||:|:|  : ||    ||| ||:   | |:||
Db        358 GAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVAA 416

Qy        423 NPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGA 475
              ||| ||:|||:        ::   | :  ||: ::||::|||    |||::|||||||| |
Db        417 NPQLDREGHVVPGYEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYLERCDGGV 475

Qy        476 FTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGVF 534
              |    | :|| :  :::||:||    ||  ||||| ||| | |:|  | |  ||||
Db        476 IVGRQEMDVFRYVADSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGVF 535

Qy        535 GTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDRG 594
              ||||| ||||| ||||| :  |::  | :|  :  ||:|||| ||:  :||  : :|:
Db        536 GTMNSAYSDCDVLGNYAAFSALKR-ADGSENTRTIMQETYRAATERVMAELEALQYVDQA 594

Qy        595 APCSSEGLSSVIVDHPTFRRILDTLR 620
              | :  | ::|         ::: ::
Db        595 VPTALGRLETIIGTREALHTVVNNIK 620


RESULT 6
AAW72196
ID   AAW72196 standard; Protein; 543 AA.
XX
AC   AAW72196;
XX
DT   13-JAN-1999  (first entry)
XX
DE   HSV-2 strain SB5 Contig ID 15 ORF#31 protein.
XX
KW   HSV-2 strain SB5; immunological response induction; therapy;
KW   antiviral identification; viral protein inhibitor.
XX
OS   Herpes simplex virus type 2.
XX
PN   WO9820016-A1.
XX
PD   14-MAY-1998.
XX
PF   31-OCT-1997;   97WO-US20016.
XX
PR   09-JUN-1997;   97US-0049018.
PR   04-NOV-1996;   96US-0030279.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Chan JY,  Dabrowski-Amaral CE,  Delvecchio AM,  Dillon SB;
PI   Esser KM,  Leary JJ;
XX
DR   WPI; 1998-286847/25.
DR   N-PSDB; AAV62176.
XX
PT   Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT   and treatment of infection or inducing immunological response in
PT   mammal
XX
PS   Claim 10; Page 122-123; 748pp; English.
XX
CC   This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC   sequence of the invention. This sequence was isolated from a HSV-2 strain
CC   SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
CC   Based on homology, this sequence is a major DNA-binding protein.
CC   The proteins can be used for the treatment or prevention of disease, to
CC   induce an immunological response in a mammal or to identify inhibitors,
CC   activators or novel antivirals. Antagonists of the proteins can be used
CC   to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC   it can also be used to induce an immunological response in a mammal.
XX
SQ   Sequence   543 AA;

  Query Match             21.9%;  Score 1380.5;  DB 19;  Length 543;
  Best Local Similarity   48.7%;  Pred. No. 2.7e-125;
  Matches  271;  Conservative  101;  Mismatches  165;  Indels   19;  Gaps    7;

Qy        653 MALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRR 712
              |:|: |||:   ||:   | :|::||| |||||||| || || |||||||||||||||||
Db          1 MSLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRR 60

Qy        713 RFVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRV 771
              | :|||| ||:| :::|| ||||  : ||: | || |||  :|:||:|||::   :::||
Db         61 RVMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRV 120

Qy        772 KNRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPP 831
              |:||:|:|   | ||||:||:  | ||||: :||||:| | |||||||| ::|| | ||
Db        121 KSRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPP 180

Qy        832 NSKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQ 891
              |  |||||||| ||||:||  |: |:| ||| :|||: :| |||||||  |  : |||
Db        181 GSNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAM 240

Qy        892 FYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLE 951
              :|||| ||:||||| | ||||::|  |:|||   :   |  :     | :|  |:||::
Db        241 YYMANQILRYCDHSTYFINTLTAVIAGSRRPPGVQAAAAWAPQ---GGAGLEAGARALMD 297

Qy        952 KTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGK 1011
              :  |  ||: | | :|:|  |  |||||||:|||||:| |||:|||||||: | |||
Db        298 SLDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGK 357

Qy       1012 NVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYA 1071
              | |||  |||||:|::|||| ||:|  :    |  | :| :|:|||| ||| |  :::
Db        358 NACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFV 417

Qy       1072 TVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVK 1127
                |:::| | | :  :|||:| :||:|: :: |  : ::    |: : ||    ||
Db        418 ATVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEA 477

Qy       1128 ILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELF 1187
                :: :   ||:         |||        |  ::   :: :  :|:      ||    |: |
```

```
Db        478 LVSQLGAAGE-------VFNFGDFGDEDDHAASFGGLAAA--AGAAGVARKRAFHGDDPF 528

Qy       1188 DLSGIPIKHGNITMEM 1203
              | | |  ::|::|
Db        529 G-EGPPEKK-DLTLDM 542


RESULT 7
AAB53126
ID   AAB53126 standard; Protein; 1132 AA.
XX
AC   AAB53126;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Macaca mulatta rhadinovirus 17577 RRV ORF6 protein SEQ ID NO:9.
XX
KW   Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
KW   genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
KW   IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
KW   cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
KW   lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
KW   splenomegaly; hypergammaglobinulinaemia; autoimmune haemolytic anaemia.
XX
OS   Macaca mulatta rhadinovirus 17577.
XX
PN   WO200028040-A2.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26260.
XX
PR   06-NOV-1998;   98US-0107507.
PR   20-NOV-1998;   98US-0109409.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Wong SW,  Axthelm MK,  Searles RP;
XX
DR   WPI; 2000-376552/32.
XX
PT   New rhesus rhadino virus for producing non-human primate model useful
PT   for testing potential treatments and efficacy of the candidate vaccine
PT   for conditions associated with RRV infection  -
XX
PS   Claim 5; Page 123; 141pp; English.
XX
CC   The present invention describes a novel rhesus macaque rhadinovirus
CC   called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
CC   RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
CC   encoded by the genome sequence. The present invention also specifically
CC   claims the individual open reading frame (ORF) nucleotide sequences from
CC   the genome which encode the individual proteins, but these sequences are
CC   not given. A non-human animal infected with RRV can be used for testing
CC   the efficacy of drug in the treatment of condition associated with
CC   infection with RRV such as Kaposi's sarcoma, lymphoproliferative
CC   disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
CC   hypergammaglobinulinaemia or autoimmune haemolytic anaemia, by
CC   administering the drug to a immuno-compromised non-human primate
CC   preferably Rhesus macaque monkey obtained by as a result of infection
CC   by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
CC   non-human primate model for testing potential treatments for conditions
CC   associated with RRV infection. It is also useful for testing the
CC   efficacy of the candidate vaccine against RRV infection or conditions
CC   associated with its infection by administering the vaccine to the
CC   subject capable of infection with RRV, inoculating the subject with RRV
CC   and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
CC   to AAB53213 represent sequence used in the exemplification of the
CC   present invention.
XX
SQ   Sequence   1132 AA;

 Query Match             9.9%;  Score 620;  DB 21;  Length 1132;
  Best Local Similarity   22.6%;  Pred. No. 2.7e-50;
  Matches  290;  Conservative  212;  Mismatches  506;  Indels  274;  Gaps   50;

Qy          1 MENTQKTVTVPTGPLGYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTS 60
              :|: | :   | | ||||| :|  | | | | | :|:  ||:: |||| |
Db         11 LEDNQGS-RAPIGACGYVYAYSKQDFPFAEASILGNRPSGSGVFSLPILYGLTVEHEFPL 69

Qy         61 SL-AVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARER 119
              :: |      || ||    :|:|   |: | ||||       |     |   | ||
Db         70 TVKAAYKKVDTTTLA-----VKVTC--FHREVIVFHNASLFRPVFDGTGLNELCEEARAL 122

Qy        120 FGFSRC--QGPP--VDGAVETTGAEICTRLGLEPENTILYLVVTALFKE---------AV 166
              ||:::   ||| : :|      |:| :  :   |:||| |||          ||
Db        123 FGYTQFIEPGPPHSIWNPLE------CPQLP-DKDEMFLGVVVTEGFKERLWRGCLVPAV 175

Qy        167 FMCNVFLHYGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPT 226
              |          | |      ::||:  ||| |  :|:
Db        176 FQTQQ---------VQIAGRQAFKVPLYDEDLFAPHGHRM-------------------P 207

Qy        227 PFYNTGLCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQS 286
              ||:  :   ::| :  :| |||::::||||          :::: | :   :    |
Db        208 RFYHKDVSAYLYDSLFTSIAQALRLKDVTAVIHATEKQFMQDHYKIAKI---VQAKQF-- 262

Qy        287 SSSGTTTARGARRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDI 346
                                ||: |  : |    :  : |: ||    :|       | | |
Db        263 ----------------STTLPKTTDGSSHMIVDSVVAELALSYGCMF-----LECPQDA 300

Qy        347 KE------WPMFIGMEGTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEA 400
              |       ||:| | :     |:|||  ::|  |  :  :|: || ||||:|:
Db        301 CELLNYDSWPIFDGCDSPEARVNALERWSAEQAVHVAGQLFAANSVLYLTKVQKQAPRGQ 360

Qy        401 KDGGPGPSFNRFYQFAGPHLAANPQTDRDGH-----VLSSQSTGSSNTEFSVDYLALICG 455
              |  |   :| |:   |                :|      | |:   ||| | :   :||
Db        361 K--GDVNVYNSFFLQHGLGFLNEATIKENGSEAFKGVPSNALDGSSFTPY---HLAYAAS 415

Qy        456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHR 515
              |   |||:| :|::      :     :  ||  :   |:|| :|   | :|  :|
Db        416 FSPHLLAKLCYYMQFLQHHKSSTNQAFNMVHYVGTAANSEMCTLCHGNTPATCLNTLFYR 475

Qy        516 LRQRMPRFGQATRQ-PIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQT-EAAKATMQDT 573
              |: | |       |: | | ||  : :: | : |||:|  :   : |:    |  | |
Db        476 LKDRFPAVTTPQRRDPYVVTGTAGT-FNDLEILGNFASFRDREEDGNPADEHPKYTYWQL 534

Qy        574 YRATLERLFIDLEQERLLDRGAPCSSEGLS-------SVIVDHPTFRRILDTLRARIEQT 626
              :   |:|              |: |::    :: | :  :| |:   : : ::
Db        535 CQTVTEKL---------------SAIGITEDHDNHVNLITNIQSFLRVFKGIDSIVDGE 578

Qy        627 TTQFMKVLVETRDYKIREGLSEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVQDLALS 686
              :|:  ::: ::  || :  | :  :  | | |  :  | :: | ::||: |
Db        579 VMKFVNSMIK-NNFNFREHVKSVHHILQFCCNVYWQAPCAVFLNLYYKSLLWIIQDICLP 637

Qy        687 QCHCVFYGQ-----------------QVEGRNFRN----------QFQPVLRRRFVDLF 718
              :|: | |                  |  ||:              : : | |  | | |
Db        638 --YCMIYEQDNPAMGILPSEWLKMHFQTLWTNFKAACLDRGVLTGCELKIVHRDMFCDFF 695

Qy        719 -----NGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKN 773
                  : | ::   : |                    :||  : | : |::||
Db        696 DTDAGSNGLMAPFKMQVR-----------------------IARAMMVVPKSIKIKN 729

Qy        774 RVVFSGNCTNLSEAARARLV---GLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMP 830
              |::||   |  ||| ::  | |  |  |  :: |  |  | |  | |  ||
Db        730 RIIFSN--TAGSEAVQSGFVKPTGTRDTY--------VVAGPYMKFLNSLHRALF----- 774

Qy        831 PNSKSPNPQWFWTLLQRNQMPADK-LTHEEITTIAA-VKRFTEEYAAINFINLPPTCIGE 888
              |::|:   :  : | |:  | : :|: : : || :  :  | :: | :
Db        775 PDTKTAALYLWHKISQTNKTPVLKDVPDDELAELVSYVKTNSLAFEETNVLDVVPDSLMS 834

Qy        889 LAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKA 948
              |:  :   ||: |  |:  ||  :|  :  |  |  |  | | | | |
Db        835 YARIKLNGAILRACGQIQFYATTL-HCLTPVLQTIDAEEYPH-----VLGSAAIATPVAY 888

Qy        949 LLE---KTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGN-- 1003
              | |   :|      |   |:   | |    ||:: : : :|| |  ||| ||  ||
```

```
Db        889 LAEIRGRT-----ALTVQTTARQPVAATGRLRPVITVPMVVNKYTGVNGNNNVFHCGNLG 943

Qy       1004 -WSGLNGGKNVCPLFT-FDRT-------RRFIIACPRGGFICPVTGPSSGNRETTLSDQV 1054

Db        944 YFAGRGVDRNLWPESSPFKKTGVSAMLRKRHVMMTP---IIDRLIKRAAG--QTISTFEA 998

Qy       1055 RGIIVSGGAMVQ----------LAIYATVVRAVGARAQHMAFDDWLSLTDDEFLARDLEELH 1106

Db        999 ESVKRSVQALLEDKDNPNLLKSVILELIRHLGKGCQD------LSSEDVQYYLGDYCMLT 1052

Qy       1107 DQIIQTLE------TPWTVEGALEAVKILDEKTTAGDGETPTNLAFNFDSCEPSHD---T 1157

Db       1053 DEVLFTLDNIAQSGVPWTIE---DAGALIEDRQDADDLQFVDSDDIATASCQPPEEQLPT 1109

Qy       1158 TSNVLNISGSNISGSTVPGLKR 1179

Db       1110 PS----------AGALLAGKKR 1121
```

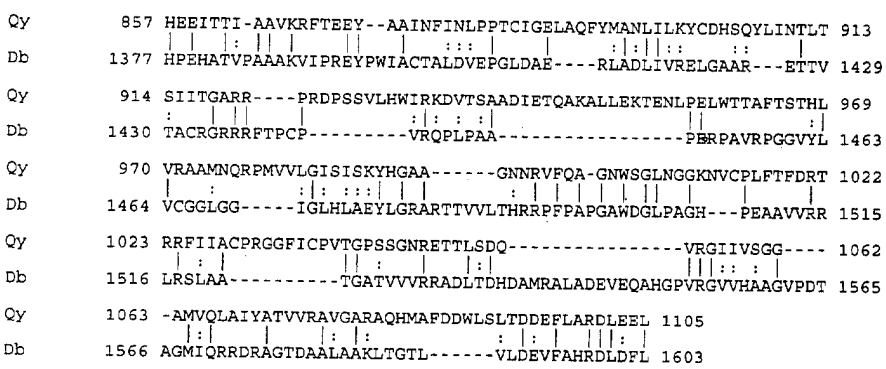
RESULT 8
AAY26935
ID AAY26935 standard; peptide; 35 AA.
XX
AC AAY26935;
XX
DT 21-DEC-1999 (first entry)
XX
DE Varicella zoster virus type 29 recombinase D35E motif.
XX
KW Recombinogenic motif; retrovirus; integrase; invertebrate; transposase;
KW consensus; vaccine; gene therapy; allergy; autoimmune disease; cancer;
KW cardiovascular disease; graft rejection; hematopoietic disorder;
KW immunosuppressive disorder; immunoproliferative disease; jaundice;
KW immunodeficiency disease; infectious disease; inflammatory disease;
KW septic shock; metabolic defect.
XX
OS Varicella zoster virus.
XX
PN US5959074-A.
XX
PD 28-SEP-1999.
XX
PF 28-FEB-1997; 97US-0807332.
XX
PR 01-MAR-1996; 96US-0012616.
PR 02-AUG-1996; 96US-0023064.
XX
PA (NAJE-) NAT JEWISH MEDICAL & RES CENT.
XX
PI Gelfand EW, Dreyfus DH;
XX
DR WPI; 1999-561074/47.
XX
PT Peptides derived from a Herpes virus recombinogenic motif useful for the
PT production of vaccines and in gene therapy -
XX
PS Claim 1; Column 37-38; 29pp; English.
XX
CC The invention relates to novel recombinogenic motifs found in
CC (retro)viral integrases which have similarity to invertebrate
CC transposase molecules. The peptides, or their homologues, have the
CC following identifying characteristics: (a) an initial and terminal amino
CC acid comprising Asp or Glu, separated by 34-39 amino acids, especially
CC 35 amino acids (the D35E motif); (b) a high probability ($p < 0.05$) of
CC alignment with a consensus sequence (AAY26941), as determined by:
CC (1) aligning the initial and terminal residues of the homologue sequence
CC with those of the consensus; (2) aligning the intervening residues of
CC the homologue with those of the consensus by maintaining the spacing of
CC the consensus while, if necessary, altering (by the insertion of spaces
CC or the deletion of residues) the spacing of the homologue; and
CC (3) generating a distribution of 1500 random sequences of amino acids
CC between the initial and terminal residues and identifying the percentage
CC of aligned residues. The recombinogenic motifs may be useful in the
CC production of vaccines and in gene therapy to treat allergies,
CC autoimmune diseases, cancers, cardiovascular diseases, graft rejection,
CC hematopoietic disorders, immunosuppressive disorders,
CC immunoproliferative diseases, immunodeficiency diseases, infectious
CC diseases, inflammatory diseases, jaundice, septic shock and other
CC immunological, genetic or metabolic defects.
XX
SQ Sequence 35 AA;

  Query Match 2.9%; Score 183; DB 20; Length 35;
  Best Local Similarity 100.0%; Pred. No. 3.5e-10;
  Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        545 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 579
              |||||||||||||||||||||||||||||||||||
Db          1 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 35
```

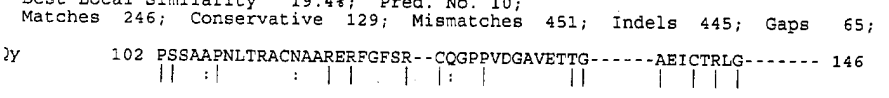
RESULT 9
AAB73931
ID AAB73931 standard; peptide; 35 AA.
XX
AC AAB73931;
XX
DT 29-MAY-2001 (first entry)
XX
DE Varicella zoster VZV29 protein recombinogenic sequence.
XX
KW Varicella zoster virus; VZV29; herpesvirus; D35E; anti-viral;
KW immunosuppressant; immunostimulant; cytostatic; anti-allergic; cardiant;
KW anti-inflammation; gene therapy; infection; herpes virus replication;
KW V(D)J recombination; retroviral integrase;
KW immunoglobulin class switching; allergy; autoimmune disease; cancer;
KW cardiovascular disease; graft rejection; haematopoietic disorder;
KW immunological disorder.
XX
OS Varicella-zoster virus.
XX
PN US6187584-B1.
XX
PD 13-FEB-2001.
XX
PF 23-JUN-1999; 99US-0338876.
XX
PR 01-MAR-1996; 96US-0012616.
PR 02-AUG-1996; 96US-0023064.
PR 28-FEB-1997; 97US-0807332.
XX
PA (NAJE-) NAT JEWISH MEDICAL & RES CENT.
XX
PI Dreyfus DH, Gelfand EW;
XX
DR WPI; 2001-256131/26.
XX
PT New nucleic acid molecules encoding peptides with transposase
PT activities, useful as immunosuppressants, anti-viral agents, or vectors
PT for gene therapy of diseases, e.g. allergies, autoimmune diseases,
PT cancers or graft rejection -
XX
PS Claim 1; Column 25; 31pp; English.
XX
CC The present sequence is provided in a specification relating to a new
CC isolated nucleic acid molecule comprising a nucleic acid sequence
CC encoding a peptide consisting of one of ten amino acid sequences
CC defined in the specification or homologues of these sequences.
CC The nucleic acid molecules are useful for producing peptides
CC important in the regulation and function of herpes virus replication,
CC V(D)J recombination, retroviral integrase function or immunoglobulin
CC class switching. The isolated nucleic acid molecules and the peptides
CC they encode are useful as immunosuppressants, anti-viral agents or
CC vectors for gene therapy. These are particularly useful in protecting
CC humans from a variety of diseases, e.g. allergies, autoimmune diseases,
CC cancers, cardiovascular diseases, graft rejection, haematopoietic

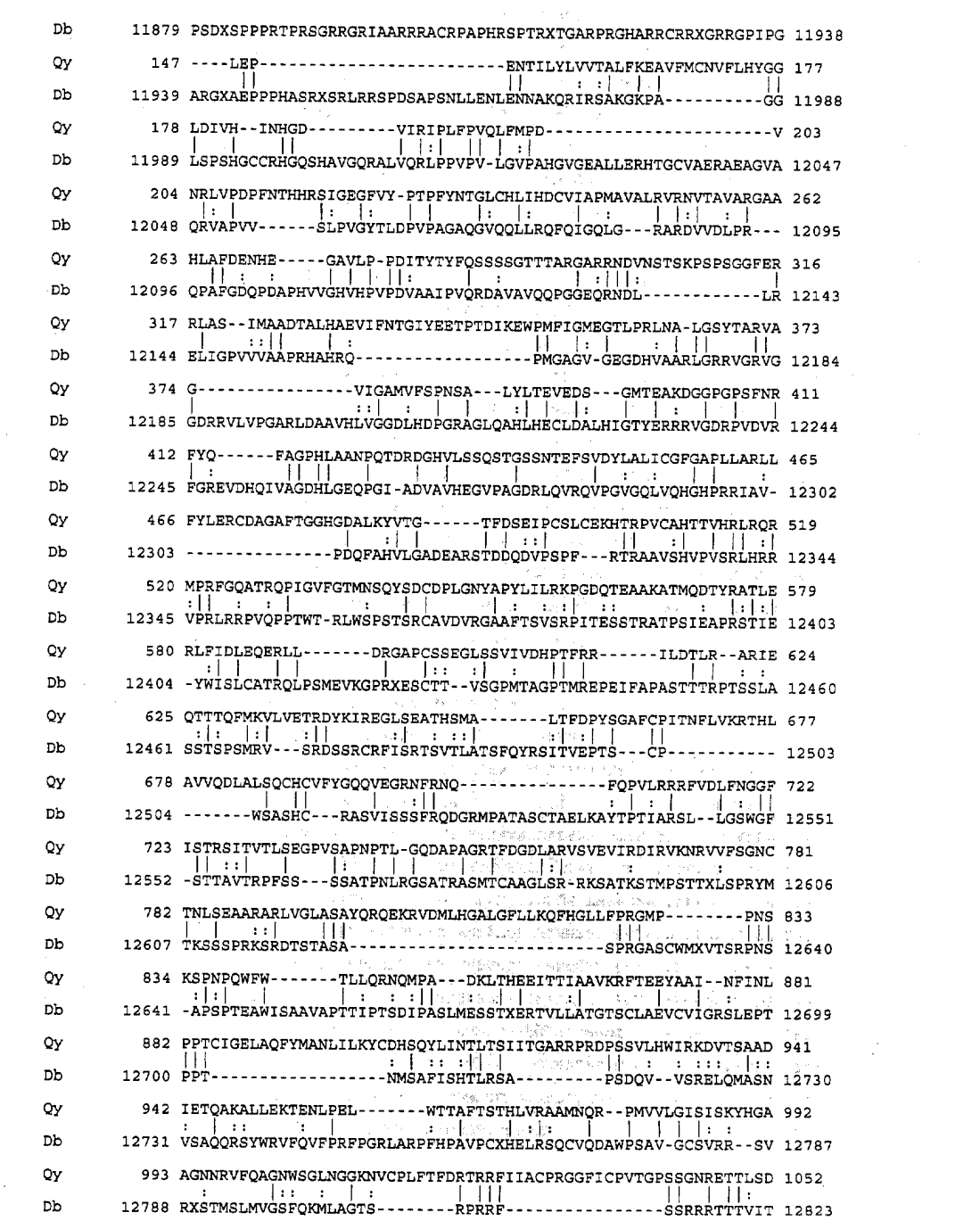
```
CC   disorders, immunodeficiency diseases, immunoproliferative diseases,
CC   immunosuppressive disorders, infectious diseases, inflammatory diseases,
CC   jaundice, septic shock, or other immunological defects, as well as other
CC   genetic or metabolic defects.
XX
SQ   Sequence   35 AA;

  Query Match             2.9%;  Score 183;  DB 22;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-10;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        545 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 579
              |||||||||||||||||||||||||||||||||||
Db          1 DPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLE 35


RESULT 10
AAB07562
ID   AAB07562 standard; Protein; 1841 AA.
XX
AC   AAB07562;
XX
DT   20-OCT-2000  (first entry)
XX
DE   Protein encoded by the bleomycin (BLM) gene cluster ORF30.
XX
KW   BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW   bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW   thiazoline; bithiazoline; microbial metabolite; sugar.
XX
OS   Streptomyces verticillus.
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00445.
XX
PR   06-JAN-1999;   99US-0115435.
PR   05-FEB-1999;   99US-0118848.
PR   05-JAN-2000; 2000US-0477962.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Shen B,  Du L,  Sanchez C,  Chen M,  Edwards DJ;
XX
DR   WPI; 2000-465974/40.
DR   N-PSDB; AAA58471.
XX
PT   New bleomycin gene cluster components useful for peptide and/or
PT   polyketide metabolites, especially bleomycin, production and for
PT   chemically modifying biological molecules -
XX
PS   Disclosure; Page 97-137; 162pp; English.
XX
CC   AAB07556-78 represent proteins encoded by open reading frames (ORFs)
CC   8 to 30 of the BLM (Bleomycin) gene cluster. The proteins encoded
CC   by the gene cluster are useful for producing peptides and/or polyketide
CC   metabolites, especially bleomycin or bleomycin analogues. They are
CC   also useful for chemically modifying biological molecules to produce
CC   branched methyl groups, and for coupling amino acids and fatty
CC   acids. They may be reacted with an apo-carrier protein and coenzyme A
CC   to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC   domains can be used individually or collectively to produce
CC   thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC   microbial metabolites. The BLM gene cluster may also be used to produce
CC   sugars.
XX
SQ   Sequence   1841 AA;

Query Match             2.5%;  Score 156.5;  DB 21;  Length 1841;
Best Local Similarity   19.6%;  Pred. No. 0.00015;
Matches  268;  Conservative  133;  Mismatches  470;  Indels  493;  Gaps   67;

Qy         14 PLGYVYACRVEDLDLEEISFLAARST-DSDLALLPLMRNLTVEKTFTSSLAVV------- 65
              |:  | : |  :   | :  ||| | :    |  |   |  : |   |||
Db        461 PMPLVVSARTREALAEAVRDLAAWSAPEPGTRLADLAATLAGRRAFPYRAAVVCHDLPEA 520

Qy         66 ----SGAR-TTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERF 120
                  |||  | | |                 :||:|  |:  ||      |   | | |
Db        521 ARLLGGARGETALPGR------------EAVFLF-PGQGTLPPDTGRGLYADVPAFRAHF 567

Qy        121 GFSRCQGPPVDGAVE---TTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGG 177
                  |   |    | ::   ||  ::|       |||     :  : :|
Db        568 ----------DACAEGFAPLGTDLHAALGAPADDT--RAAQPALFAVEYALARTLMDWGV 615

Qy        178 LDIVHINH--GDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCH 235
                :|  |: :  |  | |:||  ||          | ::  |:   |      |
Db        616 RPAAMLGHSLGEYVAATLAGV-LSLPDALTLVRARAEAQH-TMPPGRMLAVPLTPDDLRP 673

Qy        236 LIHD-----------CV-------IAPMAVALRVRNVTAVARGAAHLAFDENHEGAV-- 274
                |:                ||      :| :   |  | | |     || ||   | ||
Db        674 LLPPEVEFSAFNAPGRCVVGGPPEPVAELRARLARRGVPAAELATAH-AF---HSAAVEP 729

Qy        275 -------------LPPDITYTYFQSSSSG------TTT-------------ARGARR- 299
                          ||  : |   || :|       ||                | | ||
Db        730 LLDGFRGVLEGVRLRPPRLRYV---SSLTGDWADAAVTTPAYWLAHLRRPVRFADGLRRC 786

Qy        300 NDVN-----STSKPSPSGGFERRLAS----------IMAADTA---LHA-EVIFNTGIY- 339
                |:     |  :   | ||| |                : | | |   ||  |:: :|
Db        787 LDLGPVALVETGPRAGLTGLARRAAGPGEPPYTVRCLAAPDEAASLTHAVAVLWRSGCAV 846

Qy        340 --------------------------------EETPTDIK----------EWPMFIGME 356
                                             |  |||:          : |: :
Db        847 DWTAFHRPGRPRRTTVPGYPPQRVRHWIDAPDESEPTDLATALRAELRTDGDPPLAVDQR 906

Qy        357 -GTLPRLNALGSYTARVAGVIGAMVFSPNSALYL-TEVEDSGMTEAKDGGPGPSFNRFYQ 414
                |  || | :  ||                  || | ||:|:        | |:||
Db        907 PGLRTGLNRLCAALARD--------------YLATGVEASGVL--------PGFHRFLD 943

Qy        415 FAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAG 474
                :    |: |  | |   :::: | :                    |  : |:  |  | |
Db        944 YLRTLAASAPAAD-DAGTIAAEITAAH----------------PSFSGLVDLLRHCAQG 985

Qy        475 --------------AFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRM 520
                           :  | || |:   |                    |  ||  |:
Db        986 YPRALSTPGAALDVLYPAGSGDLLRRTLG------------------EGTADHRATGRL 1026

Qy        521 PRFG-------QATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDT 573
                |          | |:|      : :          |:  |: | ||          |
Db       1027 TRLAGSLLDRLAADREPGRPLRVLEAGAG----AGSLTQALVTRAPG---------RLDY 1073

Qy        574 YRATLERLFI-----------DLEQERLLDRGAPCSSEGLS----------SVIVDHPT 611
                :  : | |:                |  : |:||       :|  :          |:  |
Db       1074 HATDISRHFVTALGREAARRGLDFVRARVLDIARDPGEQGFAGERFDVVCGLDVVHATPD 1133

Qy        612 FRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDPYSGAFCPIT--- 668
                |  | ||: :    |     |:||                        ||:    :|
Db       1134 LRTTLGHLRSLMAPDGT---LALIET----------------TADDPWLTMIWGLTDGW 1173

Qy        669 -NFLVKRTH----------LAVVQDLALSQCHCVFYGQQVEGRNFRNQ------FQPVL 710
                 :  :|||           |  :| | :      |||                |  : :
Db       1174 WHHTDRRTHGPLLDAAGWRALLAGEDFATADVIVPPDGPQDAALLLARQTPRPAAAAPSV 1233

Qy        711 RRRFVDLF----------------NGGFI------STRSITVTLSEGPVSAPNPTLGQD 747
                :| |  :                || :      : :::    |   :  | |:
Db       1234 GKRDVGTWCYARGWRHAAPADPAPLTGGCLLLGDGDTAKAVASRLE--ALGVPVTTVGGG 1291

Qy        748 APAG----RTFDGDLARVSVEV--IRDIRVKNRVVFSGNCTNLSEAARARLVGLASAYQR 801
                | |    |   |    |::|::  :||  :  |  :          :||   |: | |:
Db       1292 RPPGPERYRELVGPATRLAVDLWPLRDASHRGRAAGAAGVRTAQDAALHNLLHLARAFGA 1351

Qy        802 QEKR-----VDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQRNQMPADKLT 856
                |:|     | :  ||       | :|                            | |
Db       1352 LEERHPARVVTVTTGA--------HDVL-------------------------GDDLA 1376
```



RESULT 11
ABP76682
ID ABP76682 standard; Protein; 19938 AA.
XX
AC ABP76682;
XX
DT 26-FEB-2003 (first entry)
XX
DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.
XX
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX
OS Streptomyces viridochromogenes.
XX
PN WO200268436-A1.
XX
PD 06-SEP-2002.
XX
PF 24-AUG-2001; 2001WO-EP09815.
XX
PR 25-FEB-2001; 2001DE-1009166.
XX
PA (COMB-) COMBINATURE BIOPHARM AG.
XX
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX
DR WPI; 2003-018650/01.
DR N-PSDB; ABZ37516.
XX
PT New avilamycin derivatives, useful for treatment of infections, and
PT nucleic acid encoding avilamycin synthesis enzymes -
XX
PS Example 1; Page 68-301; 319pp; German.
XX
CC The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by Staphylococcus
CC aureus. (I) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster
CC (ABZ37515-ABZ37516).
XX
SQ Sequence 19938 AA;

 Query Match 2.0%; Score 125.5; DB 24; Length 19938;
 Best Local Similarity 19.4%; Pred. No. 10;
 Matches 246; Conservative 129; Mismatches 451; Indels 445; Gaps 65;



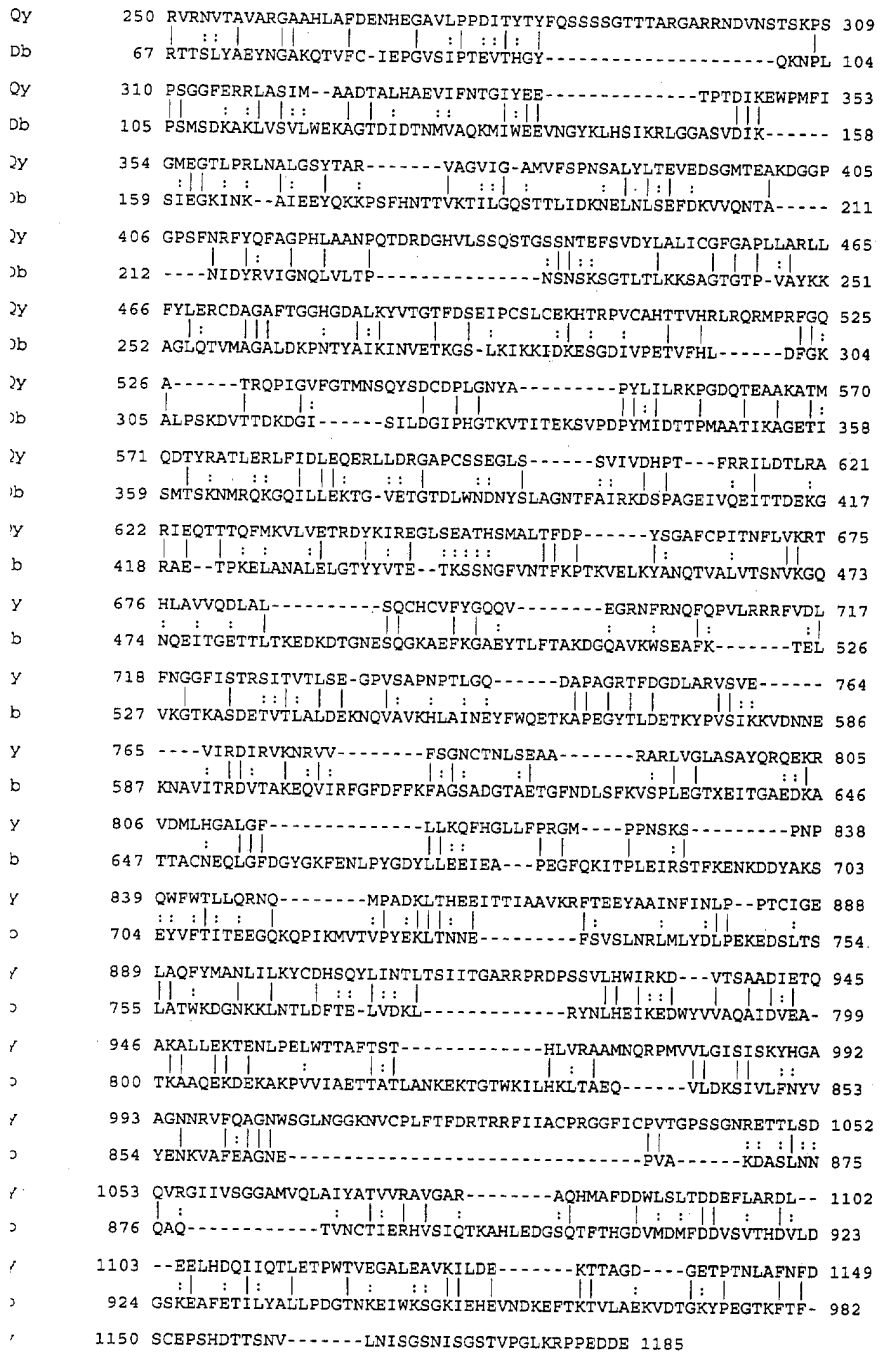


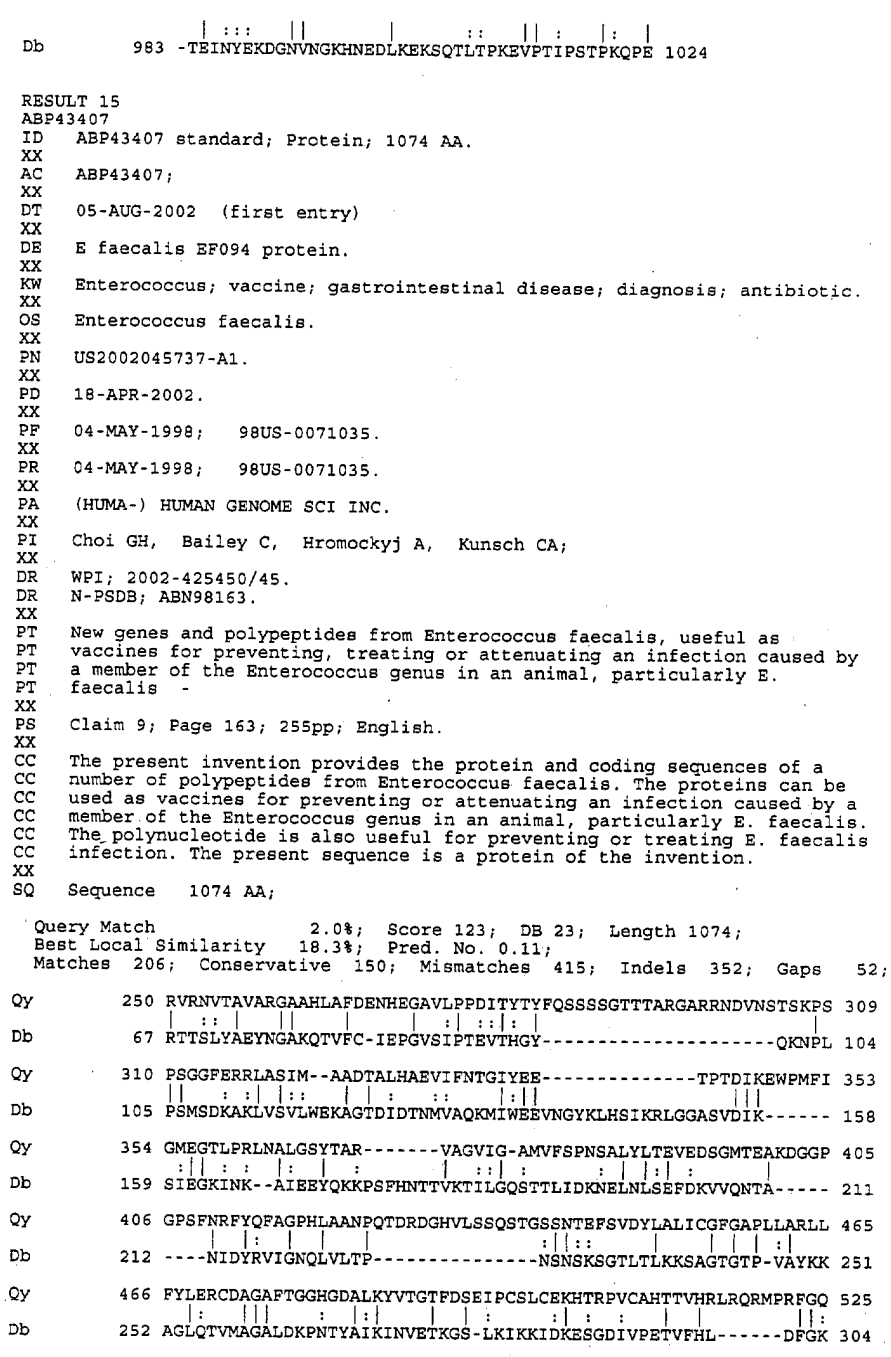
```
Qy       1053 QVRGIIVSGGAMVQLAIYATVVRAVGARAQHMAFDDWLSLTDDEPLARDLEELHDQIIQT 1112
Db      12824 ESSPCSTSG---------RRTCRAVGASRISCAI---CSRTSSS---------------- 12855

Qy       1113 LETPWTVEGALEAVKILDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGS 1172
Db      12856 ----------------TRERSSSGD-------------RPANDTTSASASPPGSSVSGA 12885

Qy       1173 ---TVPGLKRP 1180
Db      12886 AGCSTTGXRRP 12896


RESULT 12
AAB73958
ID   AAB73958 standard; Protein; 3210 AA.
XX
AC   AAB73958;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Mycelia sterilia cyclic depsipeptide synthase.
XX
KW   Mycelia sterilia; cyclic depsipeptide synthase; anthelmintic;
KW   PF1022.
XX
OS   Mycelia sterilia.
XX
PN   WO200118179-A1.
XX
PD   15-MAR-2001.
XX
PF   07-SEP-2000; 2000WO-JP06103.
XX
PR   07-SEP-1999;   99JP-0253040.
PR   06-APR-2000; 2000JP-0104291.
XX
PA   (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX
PI   Midoh N,  Okakura K,  Miyamoto K,  Watanabe M,  Yanai K,  Yasutake T;
PI   Aihara S,  Futamura T,  Kleinkauf H,  Murakami T;
XX
DR   WPI; 2001-265970/27.
DR   N-PSDB; AAF79702.
XX
PT   Novel cyclic depsipeptide synthase and gene encoding it for efficient
PT   production of anthelmintic substance PF1022  -
XX
PS   Claim 1; Page 59-80; 92pp; Japanese.
XX
CC   The present sequence is a cyclic depsipeptide synthase from Mycelia
CC   sterilia. The present protein, or a protein containing an addition,
CC   deletion and/or substitution of one or more amino acid residues is useful
CC   for the efficient production of the anthelmintic PF1022 (cyclo(D-lactyl-
CC   L-N-methylleucyl-D-3-phenyllactyl-L-N-methylleucyl-D-lactyl-L-N-
CC   methylleucyl-D-3-phenyllactyl-L-N-methylleucyl)).
XX
SQ   Sequence   3210 AA;

 Query Match             2.0%;  Score 123.5;  DB 22;  Length 3210;
 Best Local Similarity   18.7%;  Pred. No. 0.66;
 Matches  111;  Conservative   79;  Mismatches  227;  Indels  177;  Gaps   24;

Qy        298 RRNDVNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGME- 356
Db       2550 RRAQIAPKSQAAPA-------KPVKQVDPFVNLEAIL-----------CEEFAEVLGMEV 2591

Qy        357 GTLPRLNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFA 416
Db       2592 GVNDHFFQLGGHSLLATKLVARLSRRLNGRVSVRDVFDQ---------------------- 2630

Qy        417 GPHLAANPQTDRDGHVLSS----------QSTGSSNTEFSVDYLALICGFGAPLLARLL 465
Db       2631 -PVISDLAVTLRQGLTLENAIPATPDSGYWEQTMSAPTTPSDDMEAVLCKEFADVLG--- 2686

Qy        466 FYLERCDAGAFTGGHG-------------DALKYVTGTPDSEIPCSLCEKHTRPVCAHT 511
Db       2687 VEVSATDSFFDLGGHSLMATKLAARISRRLDVPVSIKDIFDHSVPLNLARK--------- 2737

Qy        512 TVHRLRQRMPRFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLIL----------RKPGD 561
Db       2738 ----IRLTQAKGHEATN---GV-----------QIANDAPFQLISVEDPEIFVQREIAP 2778

Qy        562 QTEAAKATMQDTYRAT-LERLF---------------IDLEQERLLDRGAPCSS--EGL 602
Db       2779 QLQCSPETILDVYPATQMQRVFLLNPVTGKPRSPTPFHIDFPPD------ADCASLMRAC 2832

Qy        603 SSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHS---------- 652
Db       2833 ASLAKHFDIFRTVFLEARGELYQVVLKHVDVPIEM--LQTEENINSATRSFLDVDAEKPI 2890

Qy        653 --------MALTFDPYSGAFCPITNFLVKRTHLAVVQDLALS----QCHCVFYGQQV-EG 699
Db       2891 RLGQPLIRIAILEKPGS------TLRVILRLSHALYDGLSLEHILHSLHILFFGGSLPPP 2944

Qy        700 RNFRNQFQPVL--RRRFVDLFNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGD 757
Db       2945 PKFAGYMQHVASSRREGYDFWR-SVLRDSSMTVIKGNNNTTPPPPPQQQSTPSGAHHASK 3003

Qy        758 LARVSVEVIRDIRVKNRVVFSGNCT-------NLSEAARARLV----GLASAYQ 800
Db       3004 VVTIPTQANTDSRITRATIFTTACALMLAKEDNSSDVVFGRTVSGRQGLPLAHQ 3057


RESULT 13
AAY00206
ID   AAY00206 standard; Protein; 1074 AA.
XX
AC   AAY00206;
XX
DT   20-APR-1999  (first entry)
XX
DE   Enterococcus faecalis protein EF102.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C,  Choi GH,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   N-PSDB; AAX20196.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 9; Page 197-198; 301pp; English.
XX
CC   The present sequence represents a protein isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
```

```
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
XX
SQ   Sequence   1074 AA;

  Query Match             2.0%;  Score 123;  DB 20;  Length 1074;
  Best Local Similarity   18.3%;  Pred. No. 0.11;
  Matches  206;  Conservative  150;  Mismatches  415;  Indels  352;  Gaps   52;

Qy       250 RVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRNDVNSTSKPS 309
Db        67 RTTSLYAEYNGAKQTVFC-IEPGVSIPTEVTHGY--------------------QKNPL 104

Qy       310 PSGGFERRLASIM--AADTALHAEVIFNTGIYEE-------------TPTDIKEWPMFI 353
Db       105 PSMSDKAKLVSVLWEKAGTDIDTNMVAQKMIWEEVNGYKLHSIKRLGGASVDIK------ 158

Qy       354 GMEGTLPRLNALGSYTAR-------VAGVIG-AMVFSPNSALYLTEVEDSGMTEAKDGGP 405
Db       159 SIEGKINK--AIEEYQKKPSFHNTTVKTILGQSTTLIDKNELNLSEFDKVVQNTA----- 211

Qy       406 GPSFNRFYQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAPLLARLL 465
Db       212 ----NIDYRVIGNQLVLTP--------------NSNSKSGTLTLKKSAGTGTP-VAYKK 251

Qy       466 FYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQ 525
Db       252 AGLQTVMAGALDKPNTYAIKINVETKGS-LKIKKIDKESGDIVPETVFHL------DFGK 304

Qy       526 A------TRQPIGVFGTMNSQYSDCDPLGNYA---------PYLILRKPGDQTEAAKATM 570
Db       305 ALPSKDVTTDKDGI------SILDGIPHGTKVTITEKSVPDPYMIDTTPMAATIKAGETI 358

Qy       571 QDTYRATLERLFIDLEQERLLDRGAPCSSEGLS------SVIVDHPT---FRRILDTLRA 621
Db       359 SMTSKNMRQKGQILLEKTG-VETGTDLWNDNYSLAGNTFAIRKDSPAGEIVQEITTDEKG 417

Qy       622 RIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDP------YSGAFCPITNFLVKRT 675
Db       418 RAE--TPKELANALELGTYYVTE--TKSSNGFVNTFKPTKVELKYANQTVALVTSNVKGQ 473

Qy       676 HLAVVQDLAL----------SQCHCVFYGQQV--------EGRNFRNQFQPVLRRRFVDL 717
Db       474 NQEITGETTLTKEDKDTGNESQGKAEFKGAEYTLFTAKDGQAVKWSEAFK-------TEL 526

Qy       718 FNGGFISTRSITVTLSE-GPVSAPNPTLGQ------DAPAGRTFDGDLARVSVE------ 764
Db       527 VKGTKASDETVTLALDEKNQVAVKHLAINEYFWQETKAPEGYTLDETKYPVSIKKVDNNE 586

Qy       765 ----VIRDIRVKNRVV--------FSGNCTNLSEAA-------RARLVGLASAYQRQEKR 805
Db       587 KNAVITRDVTAKEQVIRFGFDFFKFAGSADGTAETGFNDLSFKVSPLEGTXEITGAEDKA 646

Qy       806 VDMLHGALGF--------------LLKQFHGLLFPRGM----PPNSKS---------PNP 838
Db       647 TTACNEQLGFDGYGKFENLPYGDYLLEEIEA---PEGFQKITPLEIRSTFKENKDDYAKS 703

Qy       839 QWFWTLLQRNQ--------MPADKLTHEEITTIAAVKRFTEEYAAINFINLP--PTCIGE 888
Db       704 EYVFTITEEGQKQPIKMVTVPYEKLTNNE---------FSVSLNRLMLYDLPEKEDSLTS 754

Qy       889 LAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKD---VTSAADIETQ 945
Db       755 LATWKDGNKKLNTLDFTE-LVDKL-------------RYNLHEIKEDWYVVAQAIDVEA- 799

Qy       946 AKALLEKTENLPELWTTAFTST------------HLVRAAMNQRPMVVLGISISKYHGA 992
Db       800 TKAAQEKDEKAKPVVIAETTATLANKEKTGTWKILHKLTAEQ------VLDKSIVLFNYV 853

Qy       993 AGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSD 1052
Db       854 YENKVAFEAGNE--------------------------------PVA------KDASLNN 875

Qy      1053 QVRGIIVSGGAMVQLAIYATVVRAVGAR--------AQHMAFDDWLSLTDDEFLARDL-- 1102
Db       876 QAQ------------TVNCTIERHVSIQTKAHLEDGSQTFTHGDVMDMFDDVSVTHDVLD 923

Qy      1103 --EELHDQIIQTLETPWTVEGALEAVKILDE-------KTTAGD----GETPTNLAFNFD 1149
Db       924 GSKEAFETILYALLPDGTNKEIWKSGKIEHEVNDKEFTKTVLAEKVDTGKYPEGTKFTF- 982

Qy      1150 SCEPSHDTTSNV-------LNISGSNISGSTVPGLKRPPEDDE 1185
Db       983 -TEINYEKDGNVNGKHNEDLKEKSQTLTPKEVPTIPSTPKQPE 1024
```

RESULT 14
AAY00188
ID   AAY00188 standard; Protein; 1074 AA.
XX
AC   AAY00188;
XX
DT   20-APR-1999 (first entry)
XX
DE   Enterococcus faecalis protein EF094.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C, Choi GH, Hromockyj A, Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   N-PSDB; AAX20178.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 9; Page 186-187; 301pp; English.
XX
CC   The present sequence represents a protein isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
XX
SQ   Sequence   1074 AA;

  Query Match   2.0%; Score 123; DB 20; Length 1074;
  Best Local Similarity   18.3%; Pred. No. 0.11;
  Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps   52;





RESULT 15
ABP43407
ID ABP43407 standard; Protein; 1074 AA.
XX
AC ABP43407;
XX
DT 05-AUG-2002 (first entry)
XX
DE E faecalis EF094 protein.
XX
KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.
XX
OS Enterococcus faecalis.
XX
PN US2002045737-A1.
XX
PD 18-APR-2002.
XX
PF 04-MAY-1998; 98US-0071035.
XX
PR 04-MAY-1998; 98US-0071035.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR WPI; 2002-425450/45.
DR N-PSDB; ABN98163.
XX
PT New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis -
XX
PS Claim 9; Page 163; 255pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of polypeptides from Enterococcus faecalis. The proteins can be
CC used as vaccines for preventing or attenuating an infection caused by a
CC member of the Enterococcus genus in an animal, particularly E. faecalis.
CC The polynucleotide is also useful for preventing or treating E. faecalis
CC infection. The present sequence is a protein of the invention.
XX
SQ Sequence 1074 AA;

Query Match 2.0%; Score 123; DB 23; Length 1074;
Best Local Similarity 18.3%; Pred. No. 0.11;
Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

```
Qy       526 A------TRQPIGVFGTMNSQYSDCDPLGNYA---------PYLILRKPGDQTEAAKATM 570
Db       305 ALPSKDVTTDKDGI------SILDGIPHGTKVTITEKSVPDPYMIDTTPMAATIKAGETI 358

Qy       571 QDTYRATLERLFIDLEQERLLDRGAPCSSEGLS------SVIVDHPT---FRRILDTLRA 621
Db       359 SMTSKNMRQKGQILLEKTG-VETGTDLWNDNYSLAGNTFAIRKDSPAGEIVQEITTDEKG 417

Qy       622 RIEQTTTQFMKVLVETRDYKIREGLSEATHSMALTFDP------YSGAFCPITNFLVKRT 675
Db       418 RAE--TPKELANALELGTYYVTE--TKSSNGFVNTFKPTKVELKYANQTVALVTSNVKGQ 473

Qy       676 HLAVVQDLAL----------SQCHCVFYGQQV--------EGRNFRNQFQPVLRRRFVDL 717
Db       474 NQEITGETTLTKEDKDTGNESQGKAEFKGAEYTLFTAKDGQAVKWSEAFK-------TEL 526

Qy       718 FNGGFISTRSITVTLSE-GPVSAPNPTLGQ------DAPAGRTFDGDLARVSVE------ 764
Db       527 VKGTKASDETVTLALDEKNQVAVKHLAINEYFWQETKAPEGYTLDETKYPVSIKKVDNNE 586

Qy       765 ----VIRDIRVKNRVV--------FSGNCTNLSEAA-------RARLVGLASAYQRQEKR 805
Db       587 KNAVITRDVTAKEQVIRFGFDFFKFAGSADGTAETGFNDLSFKVSPLEGTXEITGAEDKA 646

Qy       806 VDMLHGALGF--------------LLKQFHGLLFPRGM----PPNSKS---------PNP 838
Db       647 TTACNEQLGFDGYGKFENLPYGDYLLEEIEA---PEGFQKITPLEIRSTFKENKDDYAKS 703

Qy       839 QWFWTLLQRNQ--------MPADKLTHEEITTIAAVKRFTEEYAAINFINLP--PTCIGE 888
Db       704 EYVFTITEEGQKQPIKMVTVPYEKLTNNE---------FSVSLNRLMLYDLPEKEDSLTS 754

Qy       889 LAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKD---VTSAADIETQ 945
Db       755 LATWKDGNKKLNTLDFTE-LVDKL------------RYNLHEIKEDWYVVAQAIDVEA- 799

Qy       946 AKALLEKTENLPELWTTAFTST------------HLVRAAMNQRPMVVLGISISKYHGA 992
Db       800 TKAAQEKDEKAKPVVIAETTATLANKEKTGTWKILHKLTAEQ------VLDKSIVLFNYV 853

Qy       993 AGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSD 1052
Db       854 YENKVAFEAGNE------------------------------PVA------KDASLNN 875

Qy      1053 QVRGIIVSGGAMVQLAIYATVVRAVGAR--------AQHMAFDDWLSLTDDEFLARDL-- 1102
Db       876 QAQ-----------TVNCTIERHVSIQTKAHLEDGSQTFTHGDVMDMFDDVSVTHDVLD 923

Qy      1103 --EELHDQIIQTLETPWTVEGALEAVKILDE-------KTTAGD----GETPTNLAFNFD 1149
Db       924 GSKEAFETILYALLPDGTNKEIWKSGKIEHEVNDKEFTKTVLAEKVDTGKYPEGTKFTF- 982

Qy      1150 SCEPSHDTTSNV-------LNISGSNISGSTVPGLKRPPEDDE 1185
Db       983 -TEINYEKDGNVNGKHNEDLKEKSQTLTPKEVPTIPSTPKQPE 1024
```

Search completed: January 30, 2004, 13:12:34
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

)M protein - protein search, using sw model

!un on: January 30, 2004, 13:14:08 ; Search time 44 Seconds
(without alignments)
5682.099 Million cell updates/sec

itle: US-09-769-699-2
erfect score: 6294
equence: 1 MENTQKTVTVPTGPLGYVYA.....DELFDLSGIPIKHGNITMEM 1203

coring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 789580 seqs, 207824079 residues

otal number of hits satisfying chosen parameters: 789580

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

esult No.	Score	% Query Match	Length	DB	ID	Description
1	3054	48.5	1452	14	US-10-050-673-2	Sequence 2, Appli
2	3015.5	47.9	1196	12	US-10-200-562-200	Sequence 200, App
3	3015.5	47.9	1196	12	US-10-237-551-200	Sequence 200, App
4	3015.5	47.9	1196	12	US-10-237-551-232	Sequence 232, App
5	788	12.5	274	12	US-10-237-551-231	Sequence 231, App
6	736	11.7	248	12	US-10-237-551-230	Sequence 230, App
7	123.5	2.0	850	15	US-10-156-761-9121	Sequence 9121, Ap
8	120.5	1.9	2597	10	US-09-905-129-2	Sequence 2, Appli
9	120.5	1.9	2597	10	US-09-905-129-10	Sequence 10, Appl
10	120.5	1.9	2597	10	US-09-905-129-13	Sequence 13, Appl
11	120.5	1.9	2597	10	US-09-991-630-2	Sequence 2, Appli
12	120.5	1.9	2597	10	US-09-991-630-10	Sequence 10, Appl
13	120.5	1.9	2597	10	US-09-991-630-13	Sequence 13, Appl
14	120	1.9	5245	12	US-10-329-079-11	Sequence 11, Appl
15	118	1.9	19608	12	US-10-084-846A-8	Sequence 8, Appli
16	114.5	1.8	1496	12	US-10-021-660-125	Sequence 125, App
17	114.5	1.8	1496	12	US-10-331-496A-28	Sequence 28, Appl
18	113.5	1.8	1477	12	US-10-274-583-20	Sequence 20, Appl
19	112	1.8	6620	12	US-10-080-334-290	Sequence 290, App
20	111	1.8	5215	10	US-09-861-289-2	Sequence 2, Appli
21	111	1.8	5215	10	US-09-860-846-2	Sequence 2, Appli
22	111	1.8	5215	11	US-09-988-384B-2	Sequence 2, Appli
23	111	1.8	5215	11	US-09-836-821-2	Sequence 2, Appli
24	111	1.8	5215	12	US-10-271-889-45	Sequence 45, Appl
25	110.5	1.8	3069	10	US-09-712-363-246	Sequence 246, App
26	109.5	1.7	2630	14	US-10-077-130-2	Sequence 2, Appli
27	109.5	1.7	7968	14	US-10-077-130-5	Sequence 5, Appli
28	108	1.7	499	12	US-10-369-493-10577	Sequence 10577, A
29	108	1.7	2596	15	US-10-307-019-6	Sequence 6, Appli
30	107.5	1.7	8026	12	US-10-132-134-12	Sequence 12, Appl
31	107	1.7	905	12	US-10-369-493-2550	Sequence 2550, Ap
32	106.5	1.7	1299	15	US-10-156-761-12449	Sequence 12449, A
33	106.5	1.7	3298	12	US-10-174-677-8	Sequence 8, Appli
34	106.5	1.7	3298	12	US-10-120-801-51	Sequence 51, Appl
35	106.5	1.7	3298	15	US-10-160-758-16	Sequence 16, Appl
36	105.5	1.7	707	15	US-10-156-761-9979	Sequence 9979, Ap
37	105.5	1.7	1563	12	US-10-334-143-34	Sequence 34, Appl
38	105.5	1.7	1967	15	US-10-219-834-85	Sequence 85, Appl
39	105.5	1.7	1967	15	US-10-225-567A-575	Sequence 575, App
40	104.5	1.7	635	11	US-09-851-847-5	Sequence 5, Appli
41	104.5	1.7	1471	7	US-08-811-519A-1	Sequence 1, Appli
42	104.5	1.7	1515	12	US-10-240-154-8	Sequence 8, Appli
43	104.5	1.7	6304	15	US-10-147-026-16	Sequence 16, Appl
44	104	1.7	3192	12	US-10-132-134-10	Sequence 10, Appl
45	103	1.6	3472	12	US-10-029-120-4	Sequence 4, Appli

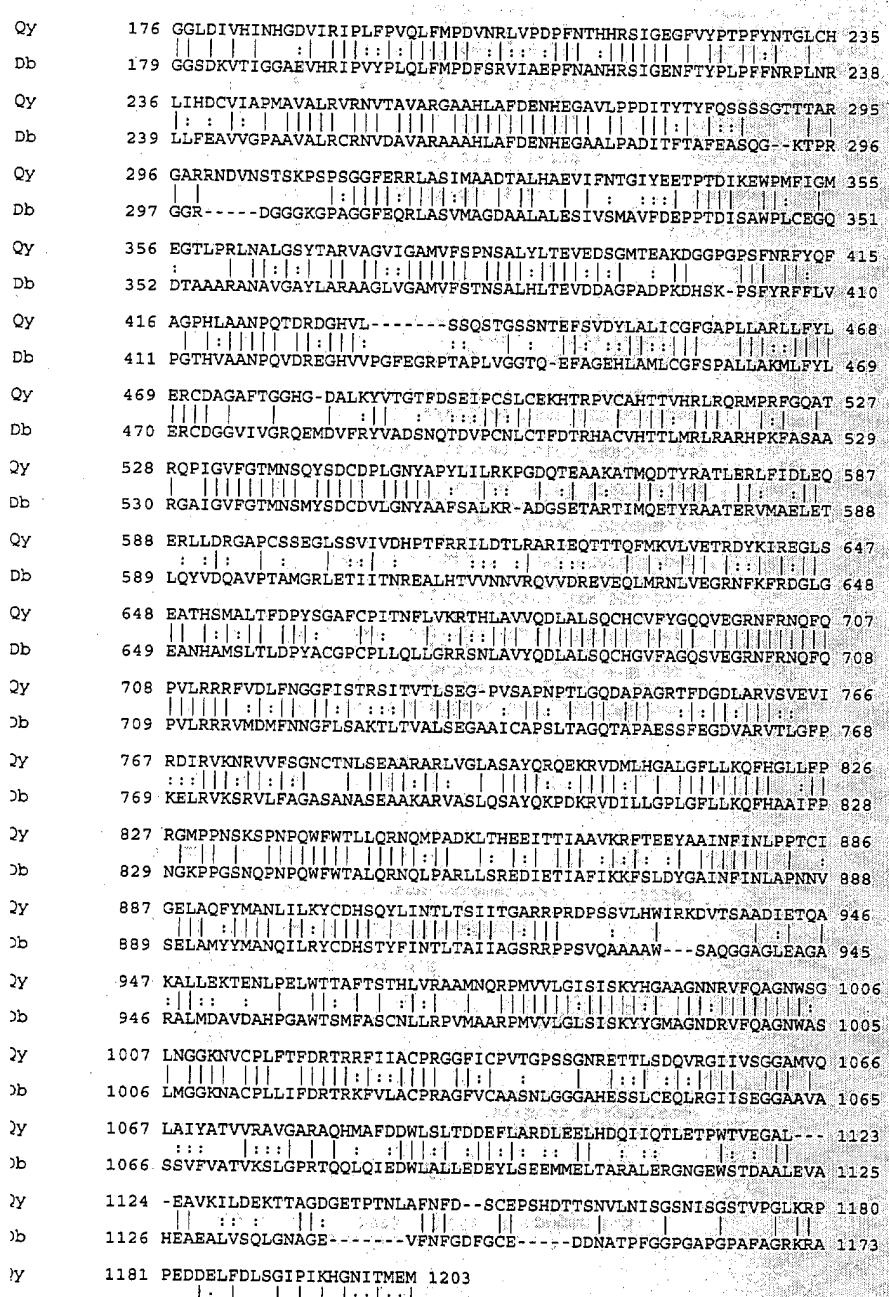
ALIGNMENTS

RESULT 1
US-10-050-673-2
; Sequence 2, Application US/10050673
; Publication No. US20020151033A1
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; TITLE OF INVENTION: Fusion Protein
; FILE REFERENCE: HU98-05
; CURRENT APPLICATION NUMBER: US/10/050,673
; CURRENT FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: US/09/127,227
; PRIOR FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-10-050-673-2

Query Match 48.5%; Score 3054; DB 14; Length 1452;
Best Local Similarity 49.8%; Pred. No. 8.3e-294;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

```
Qy          1 MENTQKTVT---VPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVE 55
              ||  || |   || ||||||  ||  | ::|  :: : ||| |:::: |::  ||||
Db          1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSGDADVAVAPLVVGLTVE 58

Qy         56 KTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNA 115
                | :::|| |:|||| |  :|||| ||:  ||:|||| : ||: |||||| |
Db         59 SGFEANVAVVVGSRTTGLGGTAVSLKLTPSHYSSSVYVFHGGRHLDPSTQAPNLTRLCER 118

Qy        116 ARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175
              ||  ||||    | |  |||| :| |||:|:  :|||| |||| : ||||
Db        119 ARRHFGFSDYTPRPGDLKHETTGEALCERLGLDPDRALLYLVVTEGFKEAVCINNTFLHL 178
```

09/769,699
Sarah Hays



RESULT 2
US-10-200-562-200
; Sequence 200, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 200
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: HSV2
US-10-200-562-200

Query Match 47.9%; Score 3015.5; DB 12; Length 1196;
Best Local Similarity 49.0%; Pred. No. 4e-290;
Matches 595; Conservative 213; Mismatches 366; Indels 41; Gaps 16;

```
Qy        714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
Db        715 VMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVK 774

Qy        773 NRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPN 832
Db        775 SRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPPG 834

Qy        833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
Db        835 SNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMY 894

Qy        893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
Db        895 YMANQILRYCDHSTYFINTLTAVIAGSRRPPSVQAAAAWAPQ---GGAGLEAGARALMDS 951

Qy        953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
Db        952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGKN 1011

Qy       1013 VCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYAT 1072
Db       1012 ACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071

Qy       1073 VVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVKI 1128
Db       1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEAL 1131

Qy       1129 LDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFD 1188
Db       1132 VSQLGAAGE-------VFNFGDFGDEDDHAASFGGLAAA--AGAAGVARKRAFHGDDPFG 1182

Qy       1189 LSGIPIKHGNITMEM 1203
Db       1183 -EGPPEKK-DLTLDM 1195
```

```
ESULT 3
S-10-237-551-200
 Sequence 200, Application US/10237551
 Publication No. US20030165820A1
 GENERAL INFORMATION:
  APPLICANT: Day, Craig H.
  APPLICANT:  Hosken, Nancy A.
  APPLICANT:  Parsons, Joseph M.
  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
  FILE REFERENCE: 210121.538C3
  CURRENT APPLICATION NUMBER: US/10/237,551
  CURRENT FILING DATE:  2002-09-06
  NUMBER OF SEQ ID NOS: 254
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 200
   LENGTH: 1196
   TYPE: PRT
   ORGANISM: HSV2
S-10-237-551-200

 Query Match           47.9%;  Score 3015.5;  DB 12;  Length 1196;
 Best Local Similarity   49.0%;  Pred. No. 4e-290;
 Matches  595;  Conservative  213;  Mismatches  366;  Indels   41;  Gaps   16;

y           4 TQKTVTVPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
b           7 TTTTVKVPPGPMGYVYGRACPAEGLEL--LSLLSARSGDADVAVAPLIVGLTVESGFEAN 64

y          62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFG 121
b          65 VAAVVGSRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQAPNLTRLCERARPHFG 124

y         122 FSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIV 181
b         125 FADYAPRPCDLKHETTGDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGGMDKV 184

Qy        182 HINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCV 241
Db        185 TIGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAV 244

Qy        242 IAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRND 301
Db        245 VGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KPQRGAR--- 299

Qy        302 VNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPR 361
Db        300 --DAGNKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPPDITTWPLLEGQETPAAR 357

Qy        362 LNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLA 421
Db        358 AGAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVA 416

Qy        422 ANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGPGAPLLARLLFYLERCDAG 474
Db        417 ANPQLDREGHVVPGYEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYLERCDGG 475

Qy        475 AFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGV 533
Db        476 VIVGRQEMDVFRYVADSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGV 535

Qy        534 FGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDR 593
Db        536 FGTMNSAYSDCDVLGNYAAFSALKR-ADGSENTRTIMQETYRAATERVMAELEALQYVDQ 594

Qy        594 GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSM 653
Db        595 AVPTALGRLETIIGNREALHTVVNNIKQLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654

Qy        654 ALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRR 713
Db        655 SLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRRR 714

Qy        714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
Db        715 VMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVK 774

Qy        773 NRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPN 832
Db        775 SRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPPG 834

Qy        833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
Db        835 SNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMY 894

Qy        893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
Db        895 YMANQILRYCDHSTYFINTLTAVIAGSRRPPSVQAAAAWAPQ---GGAGLEAGARALMDS 951

Qy        953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
Db        952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGKN 1011

Qy       1013 VCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYAT 1072
Db       1012 ACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071

Qy       1073 VVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVKI 1128
Db       1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEAL 1131

Qy       1129 LDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFD 1188
Db       1132 VSQLGAAGE-------VFNFGDFGDEDDHAASFGGLAAA--AGAAGVARKRAFHGDDPFG 1182

Qy       1189 LSGIPIKHGNITMEM 1203
Db       1183 -EGPPEKK-DLTLDM 1195
```

RESULT 4
US-10-237-551-232
; Sequence 232, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 232
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Herpes simplex virus
US-10-237-551-232

Query Match 47.9%; Score 3015.5; DB 12; Length 1196;
Best Local Similarity 49.0%; Pred. No. 4e-290;
Matches 595; Conservative 213; Mismatches 366; Indels 41; Gaps 16;

```
Qy          4 TQKTVTVPTGPLGYVY--ACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVEKTFTSS 61
              |  || || ||:|||| || | |:|  :| |:||| |:|:|: ||:  |||| | ::
Db          7 TTTTVKVPPGPMGYVYGRACPAEGLEL--LSLLSARSGDADVAVAPLIVGLTVESGFEAN 64

Qy         62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAAPNLTRACNAARERFG 121
              :| | |:||||| |  ::||| ||: |||:|||||:|: ||: |||||| |  || ||
Db         65 VAAVVGSRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQAPNLTRLCERARPHFG 124

Qy        122 FSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHYGGLDIV 181
              |:     | |   |||| :| ||||:|:  :||||:|  |:||| : | ||| ||:| |
Db        125 FADYAPRPCDLKHETTGDALCERLGLDPDRALLYLVITEGFREAVCISNTFLHLGGMDKV 184

Qy        182 HINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCV 241
              |   :| |||::|:|:|||| :|:: |||| :|||||| | || ||:|  |  |: : |
Db        185 TIGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAV 244

Qy        242 IAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQSSSSGTTTARGARRND 301
              : | ||||| ||| |||| |||||||||||||| || |||:| |::|       ||||
Db        245 VGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEASQG--KPQRGAR--- 299

Qy        302 VNSTSKPSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGMEGTLPR 361
                  |:||||:||||:|| | ||  | | :  :::| | ||  ||:  | |    |
Db        300 --DAGNKGPAGGFEQRLASVMAGDAALALESIVSMAVFDEPPPDITTWPLLEGQETPAAR 357

Qy        362 LNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNRFYQFAGPHLA 421
              |:|:| || ||::|||||| ||||:|||:|:|  : ||    ||| ||:   | |:|
Db        358 AGAVGAYLARAAGLVGAMVFSTNSALHLTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVA 416

Qy        422 ANPQTDRDGHVL-------SSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAG 474
              |||| ||:|||:       ::   | :  ||: ::||::|||   |||::|||||||| |
Db        417 ANPQLDREGHVVPGYEGRPTAPLVGGTQ-EFAGEHLAMLCGFSPALLAKMLFYLERCDGG 475

Qy        475 AFTGGHG-DALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRMPRFGQATRQPIGV 533
              .|    |  :||  :  :::||:||   ||  |||||: ||| | |:|  | |  |||
Db        476 VIVGRQEMDVFRYVADSGQTDVPCNLCTFETRHACAHTTLMRLRARHPKFASAARGAIGV 535

Qy        534 FGTMNSQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDTYRATLERLFIDLEQERLLDR 593
              |||||| ||||| ||||| :  |::  | :|  :  ||:|||  ||:  :|| : :|:
Db        536 FGTMNSAYSDCDVLGNYAAFSALKR-ADGSENTRTIMQETYRAATERVMAELEALQYVDQ 594

Qy        594 GAPCSSEGLSSVIVDHPTFRRILDTLRARIEQTTTQFMKVLVETRDYKIREGLSEATHSM 653
              |  :   | ::| :       ::: ::  :::   | |: |:| |::| |:||:|| |:|
Db        595 AVPTALGRLETIIGNREALHTVVNNIKQLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654

Qy        654 ALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFYGQQVEGRNFRNQFQPVLRRR 713
              :|: |||:   ||:  | :|::|| |||||||| || || ||||||||||||||||
Db        655 SLSLDPYTCGPCPLLQLLARRSNLAVYQDLALSQCHGVFAGQSVEGRNFRNQFQPVLRRR 714

Qy        714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVK 772
              :|||| ||:| :::|| |||| : ||: | || ||| :|:|:|:|||::  :::|||
Db        715 VMDLFNNGFLSAKTLTVALSEGAAICAPSLTAGQTAPAESSFEGDVARVTLGFPKELRVK 774

Qy        773 NRVVFSGNCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPN 832
              :||:|:|   | ||||:||:  | |||| :||||:| | |||||||| ::|| | ||
Db        775 SRVLFAGASANASEAAKARVASLQSAYQKPDKRVDILLGPLGFLLKQFHAVIFPNGKPPG 834

Qy        833 SKSPNPQWFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQF 892
              |  |||||||| |||||:||  |: |:| ||| :|||: :| ||||||| |  : ||| :
Db        835 SNQPNPQWFWTALQRNQLPARLLSREDIETIAFIKRFSLDYGAINFINLAPNNVSELAMY 894

Qy        893 YMANLILKYCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEK 952
              |||| ||:||||| | ||||||::| |:|||     :   |  :     | :|  |:||::
Db        895 YMANQILRYCDHSTYFINTLTAVIAGSRRPPSVQAAAAWAPQ---GGAGLEAGARALMDS 951

Qy        953 TENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGGKN 1012
              :  |  ||: | | :|:|   | |||||||:|||||:| |||:|||||||: | ||||
Db        952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYYGMAGNDRVFQAGNWASLLGGKN 1011

Qy       1013 VCPLFTFDRTRRFIIACPRGGFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYAT 1072
              |||  |||||:|::|||| ||:|  :     |  | :| :|:|||| |||  | :::
Db       1012 ACPLLIFDRTRKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071

Qy       1073 VVRAVGARAQHMAFDDWLSLTDDEFLARDLEELHDQIIQTLETPWTVEGAL----EAVKI 1128
              |:::| | | :  :|||:| :||:|: :: |   : ::      |: : ||    ||  :
Db       1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEEMMEFTTRALERGHGEWSTDAALEVAHEAEAL 1131

Qy       1129 LDEKTTAGDGETPTNLAFNFDSCEPSHDTTSNVLNISGSNISGSTVPGLKRPPEDDELFD 1188
              : :   ||:          |||         |  ::   :: :  :|:      ||    | |
Db       1132 VSQLGAAGE-------VFNFGDFGDEDDHAASFGGLAAA--AGAAGVARKRAFHGDDPFG 1182

Qy       1189 LSGIPIKHGNITMEM 1203
                | | |  ::|::|
Db       1183 -EGPPEKK-DLTLDM 1195
```

RESULT 5
US-10-237-551-231
; Sequence 231, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 231
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Herpes simplex virus
US-10-237-551-231

Query Match 12.5%; Score 788; DB 12; Length 274;
Best Local Similarity 56.5%; Pred. No. 2.6e-69;
Matches 153; Conservative 41; Mismatches 75; Indels 2; Gaps 1;

```
Qy         28 LEEISFLAARSTDSDLALLPLMRNLTVEKTFTSSLAVVSGARTTGLAGAGITLKLTTSHF 87
              || :| |:||| |:|:|: ||:  |||| | :::| | |:||||| |  ::||| ||:
Db          2 LELLSLLSARSGDADVAVAPLIVGLTVESGFEANVAAVVGSRTTGLGGTAVSLKLMPSHY 61

Qy         88 YPSVFVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPPVDGAVETTGAEICTRLGL 147
                 |||:|||||:|: ||: ||||||  |  ||  ||||     | |  ||||  :| ||||
Db         62 SPSVYVFHGGRHLAPSTQAPNLTRLCERARRHFGFSDYAPRPCDLKHETTGDALCERLGL 121

Qy        148 EPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLV 207
                 :|:  :||||:|  |:||| : | ||| ||:| | |   :| |||::|:|:|||| :|::
```

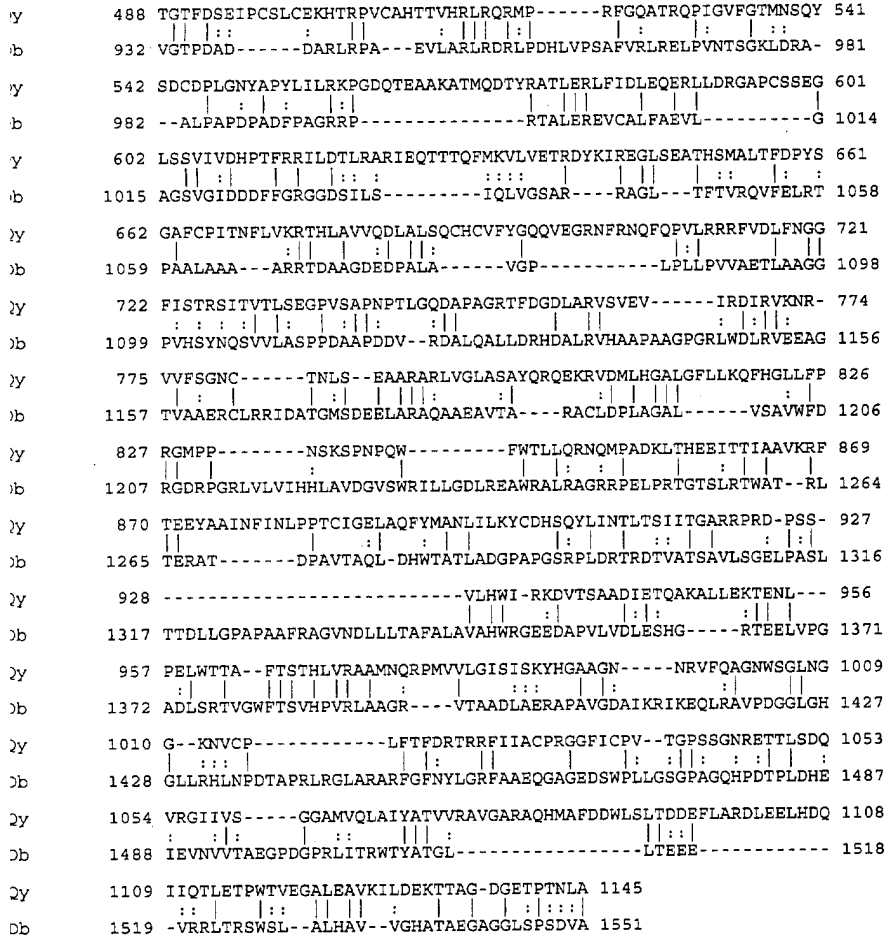
```
Db        122 DPDRALLYLVITEGFREAVCISNTFLHLGGMDKVTIGDAEVHRIPVYPLQMFMPDFSRVI 181

Qy        208 PDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAFD 267

Db        182 ADPFNCNHRSIGENFNYPLPFFNRPLARLLFEAVVGPAAVALRARNVDAVARAAAHLAFD 241

Qy        268 ENHEGAVLPPDITYTYFQSSSSGTTTARGAR 298

Db        242 ENHEGAALPADITFTAFEASQG--KPQRGAR 270


RESULT 6
US-10-237-551-230
  Sequence 230, Application US/10237551
  Publication No. US20030165820A1
  GENERAL INFORMATION:
   APPLICANT: Day, Craig H.
   APPLICANT:  Hosken, Nancy A.
   APPLICANT:  Parsons, Joseph M.
   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
   TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
   FILE REFERENCE: 210121.538C3
   CURRENT APPLICATION NUMBER: US/10/237,551
   CURRENT FILING DATE:  2002-09-06
   NUMBER OF SEQ ID NOS: 254
   SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 230
    LENGTH: 248
    TYPE: PRT
    ORGANISM: Herpes simplex virus
US-10-237-551-230

 Query Match            11.7%;  Score 736;  DB 12;  Length 248;
 Best Local Similarity   57.7%;  Pred. No. 3.3e-64;
 Matches  139;  Conservative   36;  Mismatches   66;  Indels    0;  Gaps    0;

y          47 PLMRNLTVEKTFTSSLAVVSGARTTGLAGAGITLKLTTSHFYPSVFVFHGGKHVLPSSAA 106

b           3 PLIVGLTVESGFEANVAAVVGSRTTGLGGTAVSLKLMPSHYSPSVYVFHGGRHLAPSTQA 62

y         107 PNLTRACNAARERFGFSRCQGPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAV 166

b          63 PNLTRLCERARRHFGFSDYAPRPCDLKHETTGDALCERLGLDPDRALLYLVITEGFREAV 122

y         167 FMCNVFLHYGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVPDPFNTHHRSIGEGFVYPT 226

b         123 CISNTFLHLGGMDKVTIGDAEVHRIPVYPLQMFMPDFSRVIADPFNCNHRSIGENFNYPL 182

y         227 PFYNTGLCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAFDENHEGAVLPPDITYTYFQS 286

b         183 PFFNRPLARLLFEAVVGPAAVALRARNVDAVARAAAHLAFDENHEGAALPADITFTAFEA 242

y         287 S 287

b         243 S 243


ESULT 7
S-10-156-761-9121
 Sequence 9121, Application US/10156761
 Publication No. US20030119018A1
 GENERAL INFORMATION:
  APPLICANT: OMURA, SATOSHI
  APPLICANT:  IKEDA, HARUO
  APPLICANT:  ISHIKAWA, JUN
  APPLICANT:  HORIKAWA, HIROSHI
  APPLICANT:  SHIBA, TADAYOSHI
  APPLICANT:  SAKAKI, YOSHIYUKI
  APPLICANT:  HATTORI, MASAHIRA
  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
  FILE REFERENCE: 249-262
  CURRENT APPLICATION NUMBER: US/10/156,761
  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9121
;   LENGTH: 850
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-9121

  Query Match            2.0%;  Score 123.5;  DB 15;  Length 850;
  Best Local Similarity   20.5%;  Pred. No. 0.031;
  Matches  182;  Conservative  104;  Mismatches  327;  Indels  273;  Gaps   37;

Qy        305 TSKPSPSGGFE---------RRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355

Db          3 TSRRTPSAGAEPPSRPPVRGRRAHAGQPADE--------DTGTDETVTAASADSPVGAGR 54

Qy        356 EGTLPR----------------LNALGSYTARVAGVIGAMVFSPNSALYLTEVEDSGMTE 399

Db         55 WGVRPRTVRAKIVCLLMVPVVSLLALWAY---------ATVSTAQDVARLRQVQQVDTTV 105

Qy        400 AKDGGPGPSFNRFYQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLAL----ICG 455

Db        106 RAPVAAAVAALQAERAAAVRHVIDPSAEPD----SGFRTLAARTDRAVDKLRLGGHHTVA 161

Qy        456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHR 515

Db        162 DGADLPAQVPGRLETFVSGA------EQLRSLRG----------------------AVLER 194

Qy        516 LRQRMPRFGQATRQPIGVFGTMN--SQYSDCDPLGNYAPYLILRKPGDQTEAAKATMQDT 573

Db        195 RARWDETFGQYTRTIAAAFGVGGALTGIQDAD-LGSDARVLL-----EFSRAGEALAQED 248

Qy        574 YRATLERLFIDLEQERL-LDRGAPCSSEGLS-SVIVDHPTFRRILDTLRARIEQTTTQFM 631

Db        249 AVLSSARLAGTLDGERLRLFTGAVDTRRTLTDSAVAD------LSERERAAWQGVATGRA 302

Qy        632 KVLVETRDYKI---REGLSEATHSMALTFDP------------------YSGAFCPITN 669

Db        303 YADVRTAEDKVLANRPGARRIAAAPQATWDPAHARVQEGMRTIEADAGRGVADRADPLTR 362

Qy        670 FLVKRTHLAVVQDLALSQCHCVF-----YGQQVEGRNFRNQFQPVLRRRFVD----LFNG 720

Db        363 GLLTPAGAAVLFGLAAVAASLVISVRIGRGLVIELISLRNSALEIARRKLPQAMRKLRAG 422

Qy        721 GFISTRSITVTLSEGPVSAP-NPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSG 779

Db        423 EEIDVR------AEAPPGPPAEDETGQVAEALSTVHRAALRAAVE-------------- 461

Qy        780 NCTNLSEAARARLVGLASAYQRQEKRVDMLHGALGFLLKQFHGLLFPRGMPPNSKSPNPQ 839

Db        462 ------------------------RAELASGISGV---------------------- 472

Qy        840 WFWTLLQRNQMPADKLTHEEITTIAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLIL 899

Db        473 -FVNLARRSQI----LVHRQLSLLDSMERRSED----------PNELSDL--FRLDHLTT 515

Qy        900 KYCDHSQYLINTLTSIITGA---RRPRDPSSVLHWIRKDVTSAAD---IETQ-------- 945

Db        516 RMRRHAESLI-----ILSGAAPGRAWRMPVSLTNVVRAAVSEVEDYARVEVRQLPEASVV 570

Qy        946 AKALLEKTENLPELWTTA--FTSTHLVRAAMNQRP--------MVVLGISISKYHGAAGN 995

Db        571 GAAVADLTHLMAEIVENAAQFSPPH-TRVRVTGEPVGNGYAVEVEDRGLGMGKETLAEAN 629

Qy        996 NRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIA--------------CPRGGFICPVTGP 1041

Db        630 RRIEQS----------EALDLFDSDRLGLFVVSRLAARHGIKVHLRTSPYGGTTAVVLLP 679

Qy       1042 S----SGNRETT--LSDQVRGIIVSGGAMVQLAIYATVVRAVGARA 1081
```

```
Db        680 TALLHSGTAERVPRAADTGRD-AEPAYARVAAASHQSVQQAVGRPA 724

RESULT 8
US-09-905-129-2
; Sequence 2, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
;  APPLICANT: Einat, et al
;  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 540579-2007.2
;  CURRENT APPLICATION NUMBER: US/09/905,129
;  CURRENT FILING DATE:  2001-07-13
;  PRIOR APPLICATION NUMBER: 09/802,318
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/207,821
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/084,944
;  PRIOR FILING DATE: 1998-05-11
;  PRIOR APPLICATION NUMBER: 60/085,673
;  PRIOR FILING DATE: 1998-05-15
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 2597
;   TYPE: PRT
;   ORGANISM: rattus species
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2597)
;   OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-2

  Query Match             1.9%;  Score 120.5;  DB 10;  Length 2597;
  Best Local Similarity   18.6%;  Pred. No. 0.46;
  Matches  145;  Conservative  113;  Mismatches  287;  Indels  233;  Gaps   40;

Qy        416 AGPHLAANPQTD---RDGHVLSSQSTGSSNTE----------------FSVDYLALICG 455
Db        274 SGAFLCTKPTIDPSLKSKSLVTQEDNGSASTSPQDFIEPPFGSLSLNMTXXSGNKADMVCS 333

Qy        456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV----CAHT 511
Db        334 IQKP---------SRTSPTAFTEENDYIM--LNASFSTNLVCSVDYNHIQPVWQLLALYS 382

Qy        512 TVHRLRQRMP----------RFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGD 561
Db        383 DSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEADVR-ADPFWFQQEKIVLQLNRT 441

Qy        562 QTEAAKATMQDTYRATLERLFIDLEQERL------------LDR--------GAPCSSEG 601
Db        442 ATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNNPKLERTVLVGGTIALSCPGKG 501

Qy        602 -----LSSVIVDHPTFR--------RILDTLRARIE-QTTTQFMKVL---VETRD----- 639
Db        502 DPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQMADSFDAGLYHCISTNDADADV 561

Qy        640 --YKIR--EGLSEATHSMALTFDPYSGAF----CPITNFLVKRTHLAVVQDLALSQCHCV 691
Db        562 LTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTG----------VPDASIS---WI 608

Qy        692 FYGQQVEGRNFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS--APNPT------ 743
Db        609 LPGNTVFSQPSRDR----------QILNNGTLRILQVT-PKDQGHYQCVAANPSGADFSS 657

Qy        744 ------------LGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARAR 791
Db        658 FKVSVQKKGQRMVEHDREAGGSGLGE-PNSSVSLKQPASLK----LSASALTGSEAGKQ- 711

Qy        792 LVGLASAYQRQEKRVDMLHGALG-FLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQ---R 847
Db        712 ----VSGVHRKNKHRDLIHRRRGDSTLRRFRE--HRRQLPLSARRIDPQRWAALLEKAKK 765

Qy        848 NQMPADKLTHEEITT-------IAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
Db        766 NSVP----KKQENTTVKPVPLAVPLVELTDEEKDASGMI--PP---------------- 802

Qy        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
Db        803 --DEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPQTLQ-SEHLPDFK 859

Qy        961 TTAFTSTHLVRAAM------------NQRPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
Db        860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP-SVAEIRDSA------QAGRASSQS 912

Qy       1007 ---LNGGKNVCPLFTFDRTRRFIIACPRGGFI------------CPVTG---PSSGN 1045
Db        913 AHPVTGGN----MATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSSSD 966


RESULT 9
US-09-905-129-10
; Sequence 10, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
;  APPLICANT: Einat, et al
;  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 540579-2007.2
;  CURRENT APPLICATION NUMBER: US/09/905,129
;  CURRENT FILING DATE:  2001-07-13
;  PRIOR APPLICATION NUMBER: 09/802,318
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/207,821
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/084,944
;  PRIOR FILING DATE: 1998-05-11
;  PRIOR APPLICATION NUMBER: 60/085,673
;  PRIOR FILING DATE: 1998-05-15
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;   LENGTH: 2597
;   TYPE: PRT
;   ORGANISM: Rattus sp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2597)
;   OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-10

  Query Match             1.9%;  Score 120.5;  DB 10;  Length 2597;
  Best Local Similarity   18.6%;  Pred. No. 0.46;
  Matches  145;  Conservative  113;  Mismatches  287;  Indels  233;  Gaps   40;

Qy        416 AGPHLAANPQTD---RDGHVLSSQSTGSSNTE----------------FSVDYLALICG 455
Db        274 SGAFLCTKPTIDPSLKSKSLVTQEDNGSASTSPQDFIEPPFGSLSLNMTXXSGNKADMVCS 333

Qy        456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV----CAHT 511
Db        334 IQKP---------SRTSPTAFTEENDYIM--LNASFSTNLVCSVDYNHIQPVWQLLALYS 382

Qy        512 TVHRLRQRMP----------RFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGD 561
Db        383 DSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEADVR-ADPFWFQQEKIVLQLNRT 441

Qy        562 QTEAAKATMQDTYRATLERLFIDLEQERL------------LDR--------GAPCSSEG 601
Db        442 ATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNNPKLERTVLVGGTIALSCPGKG 501

Qy        602 -----LSSVIVDHPTFR--------RILDTLRARIE-QTTTQFMKVL---VETRD----- 639
Db        502 DPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQMADSFDAGLYHCISTNDADADV 561
```

```
)y        640 --YKIR--EGLSEATHSMALTFDPYSGAF----CPITNFLVKRTHLAVVQDLALSQCHCV 691
)b        562 LTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTG----------VPDASIS---WI 608

)y        692 FYGQQVEGRNFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS--APNPT------ 743
)b        609 LPGNTVFSQPSRDR----------QILNNGTLRILQVT-PKDQGHYQCVAANPSGADFSS 657

)y        744 ------------LGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARAR 791
)b        658 FKVSVQKKGQRMVEHDREAGGSGLGE-PNSSVSLKQPASLK----LSASALTGSEAGKQ- 711

)y        792 LVGLASAYQRQEKRVDMLHGALG-FLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQ---R 847
)b        712 ----VSGVHRKNKHRDLIHRRRGDSTLRRFRE--HRRQLPLSARRIDPQRWAALLEKAKK 765

)y        848 NQMPADKLTHEEITT-------IAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
)b        766 NSVP----KKQENTTVKPVPLAVPLVELTDEEKDASGMI--PP---------------- 802

)y        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
)b        803 --DEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPQTLQ-SEHLPDFK 859

)y        961 TTAFTSTHLVRAAM-----------NQRPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
)b        860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP-SVAEIRDSA------QAGRASSQS 912

)y       1007 ---LNGGKNVCPLFTFDRTRRFIIACPRGGFI-----------CPVTG---PSSGN 1045
)b        913 AHPVTGGN----MATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSSSD 966

ESULT 10
3-09-905-129-13
 Sequence 13, Application US/09905129
 Patent No. US20020137705A1
 GENERAL INFORMATION:
  APPLICANT: Einat, et al
  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
  TITLE OF INVENTION:  AND USES THEREOF
  FILE REFERENCE: 540579-2007.2
  CURRENT APPLICATION NUMBER: US/09/905,129
  CURRENT FILING DATE:  2001-07-13
  PRIOR APPLICATION NUMBER: 09/802,318
  PRIOR FILING DATE: 2001-03-08
  PRIOR APPLICATION NUMBER: 60/207,821
  PRIOR FILING DATE: 2000-05-30
  PRIOR APPLICATION NUMBER: 60/084,944
  PRIOR FILING DATE: 1998-05-11
  PRIOR APPLICATION NUMBER: 60/085,673
  PRIOR FILING DATE: 1998-05-15
  NUMBER OF SEQ ID NOS: 25
  SOFTWARE: PatentIn version 3.0
 SEQ ID NO 13
   LENGTH: 2597
   TYPE: PRT
   ORGANISM: Rattus sp.
   FEATURE:
   NAME/KEY: misc_feature
   LOCATION: (1)..(2597)
   OTHER INFORMATION: 'x' can be any amino acid
:-09-905-129-13

Query Match             1.9%;  Score 120.5;  DB 10;  Length 2597;
Best Local Similarity   18.6%;  Pred. No. 0.46;
Matches  145;  Conservative  113;  Mismatches  287;  Indels  233;  Gaps   40;

        416 AGPHLAANPQTD---RDGHVLSSQSTGSSNTE----------------FSVDYLALICG 455
        274 SGAFLCTKPTIDPSLKSKSLVTQEDNGSASTSPQDFIEPFGSLSLNMTXXSGNKADMVCS 333

        456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV----CAHT 511

Db        334 IQKP---------SRTSPTAFTEENDYIM--LNASFSTNLVCSVDYNHIQPVWQLLALYS 382

Qy        512 TVHRLRQRMP----------RFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGD 561
Db        383 DSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEADVR-ADPFWFQQEKIVLQLNRT 441

Qy        562 QTEAAKATMQDTYRATLERLFIDLEQERL----------LDR--------GAPCSSEG 601
Db        442 ATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNNPKLERTVLVGGTIALSCPGKG 501

Qy        602 -----LSSVIVDHPTFR--------RILDTLRARIE-QTTTQFMKVL---VETRD----- 639
Db        502 DPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQMADSFDAGLYHCISTNDADADV 561

Qy        640 --YKIR--EGLSEATHSMALTFDPYSGAF----CPITNFLVKRTHLAVVQDLALSQCHCV 691
Db        562 LTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTG----------VPDASIS---WI 608

Qy        692 FYGQQVEGRNFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS--APNPT------ 743
Db        609 LPGNTVFSQPSRDR----------QILNNGTLRILQVT-PKDQGHYQCVAANPSGADFSS 657

Qy        744 ------------LGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARAR 791
Db        658 FKVSVQKKGQRMVEHDREAGGSGLGE-PNSSVSLKQPASLK----LSASALTGSEAGKQ- 711

Qy        792 LVGLASAYQRQEKRVDMLHGALG-FLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQ---R 847
Db        712 ----VSGVHRKNKHRDLIHRRRGDSTLRRFRE--HRRQLPLSARRIDPQRWAALLEKAKK 765

Qy        848 NQMPADKLTHEEITT-------IAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
Db        766 NSVP----KKQENTTVKPVPLAVPLVELTDEEKDASGMI--PP---------------- 802

Qy        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
Db        803 --DEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPQTLQ-SEHLPDFK 859

Qy        961 TTAFTSTHLVRAAM-----------NQRPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
Db        860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP-SVAEIRDSA------QAGRASSQS 912

Qy       1007 ---LNGGKNVCPLFTFDRTRRFIIACPRGGFI-----------CPVTG---PSSGN 1045
Db        913 AHPVTGGN----MATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSSSD 966

RESULT 11
US-09-991-630-2
; Sequence 2, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
;  APPLICANT: Einat, et al
;  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THEREI
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 540579-2007.3
;  CURRENT APPLICATION NUMBER: US/09/991,630
;  CURRENT FILING DATE:  2001-11-06
;  PRIOR APPLICATION NUMBER: 09/905,129
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: 09/802,318
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 09/729,485
;  PRIOR FILING DATE: 2000-12-04
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 2597
;   TYPE: PRT
;   ORGANISM: Rattus species
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (1)..(2597)
;   OTHER INFORMATION: 'x' can be any amino acid
US-09-991-630-2

  Query Match             1.9%;  Score 120.5;  DB 10;  Length 2597;
  Best Local Similarity   18.6%;  Pred. No. 0.46;
  Matches  145;  Conservative  113;  Mismatches  287;  Indels  233;  Gaps   40;

Qy        416 AGPHLAANPQTD---RDGHVLSSQSTGSSNTE----------------FSVDYLALICG 455
              :|  |    |  |    :   ::: :  ||::|                | :   ::|
Db        274 SGAFLCTKPTIDPSLKSKSLVTQEDNGSASTSPQDFIEPPFGSLSLNMTXXSGNKADMVCS 333

Qy        456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV----CAHT 511
                 |          |  |||  :   :  :  :| : : ||:   | :||      ::
Db        334 IQKP---------SRTSPTAFTEENDYIM--LNASFSTNLVCSVDYNHIQPVWQLLALYS 382

Qy        512 TVHRLRQRMP----------RFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGD 561
                : :| |           |: |   :|  :| :: :     ||        ::|:
Db        383 DSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEADVR-ADPFWFQQEKIVLQLNRT 441

Qy        562 QTEAAKATMQDTYRATLERLFIDLEQERL------------LDR--------GAPCSSEG 601
                |  :   :| :  | :     ::  |||            |:|              | :|
Db        442 ATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNNPKLERTVLVGGTIALSCPGKG 501

Qy        602 -----LSSVIVDHPTFR--------RILDTLRARIE-QTTTQFMKVL---VETRD----- 639
                  |  :: |          |||     ::| | |   |  |  : | |
Db        502 DPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQMADSFDAGLYHCISTNDADADV 561

Qy        640 --YKIR--EGLSEATHSMALTFDPYSGAF----CPITNFLVKRTHLAVVQDLALSQCHCV 691
                |:|    |   |:||   :     :|    |  |             | | ::|   :
Db        562 LTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTG----------VPDASIS---WI 608

Qy        692 FYGQQVEGRNFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS--APNPT------ 743
                |  |  :  |::           : | | :   :|    :|      | ||:
Db        609 LPGNTVFSQPSRDR----------QILNNGTLRILQVT-PKDQGHYQCVAANPSGADFSS 657

Qy        744 ------------LGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARAR 791
                         :  |  || :  |:    || : :   :|     | :    ||| :
Db        658 FKVSVQKKGQRMVEHDREAGGSGLGE-PNSSVSLKQPASLK----LSASALTGSEAGKQ- 711

Qy        792 LVGLASAYQRQEKRVDMLHGALG-FLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQ---R 847
                |   |: |  |::|   |   |::|     | :| ::: :|| : ||:   :
Db        712 ----VSGVHRKNKHRDLIHRRRGDSTLRRFRE--HRRQLPLSARRIDPQRWAALLEKAKK 765

Qy        848 NQMPADKLTHEEITT-------IAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
              | :|       :| ||      :  |:   ||  |   |  ||
Db        766 NSVP----KKQENTTVKPVPLAVPLVELTDEEKDASGMI--PP---------------- 802

Qy        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
                |    :: |  | : |    |  | |      : |  :: |      |: :|:||:
Db        803 --DEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPQTLQ-SEHLPDFK 859

Qy        961 TTAFTSTHLVRAAM----------NQRPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
              : |:   |  :|           || |:::   |:::   :|      |||  |
Db        860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP-SVAEIRDSA------QAGRASSQS 912

Qy       1007 ---LNGGKNVCPLFTFDRTRRFIIACPRGGFI----------CPVTG---PSSGN 1045
                : ||     : |:  |  :    :  :             :||   ||| :
Db        913 AHPVTGGN----MATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSSSD 966


RESULT 12
JS-09-991-630-10
: Sequence 10, Application US/09991630
: Patent No. US20020151514A1
: GENERAL INFORMATION:
:  APPLICANT: Einat, et al
:  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
:  TITLE OF INVENTION:  AND USES THEREOF
:  FILE REFERENCE: 540579-2007.3
:  CURRENT APPLICATION NUMBER: US/09/991,630
:  CURRENT FILING DATE:  2001-11-06
:  PRIOR APPLICATION NUMBER: 09/905,129
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: 09/802,318
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 09/729,485
;  PRIOR FILING DATE: 2000-12-04
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;  LENGTH: 2597
;  TYPE: PRT
;  ORGANISM: Rattus species
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(2597)
;  OTHER INFORMATION: 'x' can be any amino acid
US-09-991-630-10

  Query Match             1.9%;  Score 120.5;  DB 10;  Length 2597;
  Best Local Similarity   18.6%;  Pred. No. 0.46;
  Matches  145;  Conservative  113;  Mismatches  287;  Indels  233;  Gaps   40;

Qy        416 AGPHLAANPQTD---RDGHVLSSQSTGSSNTE----------------FSVDYLALICG 455
              :|  |    |  |    :   ::: :  ||::|                | :   ::|
Db        274 SGAFLCTKPTIDPSLKSKSLVTQEDNGSASTSPQDFIEPPFGSLSLNMTXXSGNKADMVCS 333

Qy        456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV----CAHT 511
                 |          |  |||  :   :  :  :| : : ||:   | :||      ::
Db        334 IQKP---------SRTSPTAFTEENDYIM--LNASFSTNLVCSVDYNHIQPVWQLLALYS 382

Qy        512 TVHRLRQRMP----------RFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGD 561
                : :| |           |: |   :|  :| :: :     ||        ::|:
Db        383 DSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEADVR-ADPFWFQQEKIVLQLNRT 441

Qy        562 QTEAAKATMQDTYRATLERLFIDLEQERL------------LDR--------GAPCSSEG 601
                |  :   :| :  | :     ::  |||            |:|              | :|
Db        442 ATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNNPKLERTVLVGGTIALSCPGKG 501

Qy        602 -----LSSVIVDHPTFR--------RILDTLRARIE-QTTTQFMKVL---VETRD----- 639
                  |  :: |          |||     ::| | |   |  |  : | |
Db        502 DPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQMADSFDAGLYHCISTNDADADV 561

Qy        640 --YKIR--EGLSEATHSMALTFDPYSGAF----CPITNFLVKRTHLAVVQDLALSQCHCV 691
                |:|    |   |:||   :     :|    |  |             | | ::|   :
Db        562 LTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTG----------VPDASIS---WI 608

Qy        692 FYGQQVEGRNFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS--APNPT------ 743
                |  |  :  |::           : | | :   :|    :|      | ||:
Db        609 LPGNTVFSQPSRDR----------QILNNGTLRILQVT-PKDQGHYQCVAANPSGADFSS 657

Qy        744 ------------LGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARAR 791
                         :  |  || :  |:    || : :   :|     | :    ||| :
Db        658 FKVSVQKKGQRMVEHDREAGGSGLGE-PNSSVSLKQPASLK----LSASALTGSEAGKQ- 711

Qy        792 LVGLASAYQRQEKRVDMLHGALG-FLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQ---R 847
                |   |: |  |::|   |   |::|     | :| ::: :|| : ||:   :
Db        712 ----VSGVHRKNKHRDLIHRRRGDSTLRRFRE--HRRQLPLSARRIDPQRWAALLEKAKK 765

Qy        848 NQMPADKLTHEEITT-------IAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
              | :|       :| ||      :  |:   ||  |   |  ||
Db        766 NSVP----KKQENTTVKPVPLAVPLVELTDEEKDASGMI--PP---------------- 802

Qy        901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
                |    :: |  | : |    |  | |      : |  :: |      |: :|:||:
Db        803 --DEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPQTLQ-SEHLPDFK 859

Qy        961 TTAFTSTHLVRAAM----------NQRPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
              : |:   |  :|           || |:::   |:::   :|      |||  |
Db        860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP-SVAEIRDSA------QAGRASSQS 912

Qy       1007 ---LNGGKNVCPLFTFDRTRRFIIACPRGGFI----------CPVTG---PSSGN 1045
                : ||     : |:  |  :    :  :             :||   ||| :
Db        913 AHPVTGGN----MATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSSSD 966
```

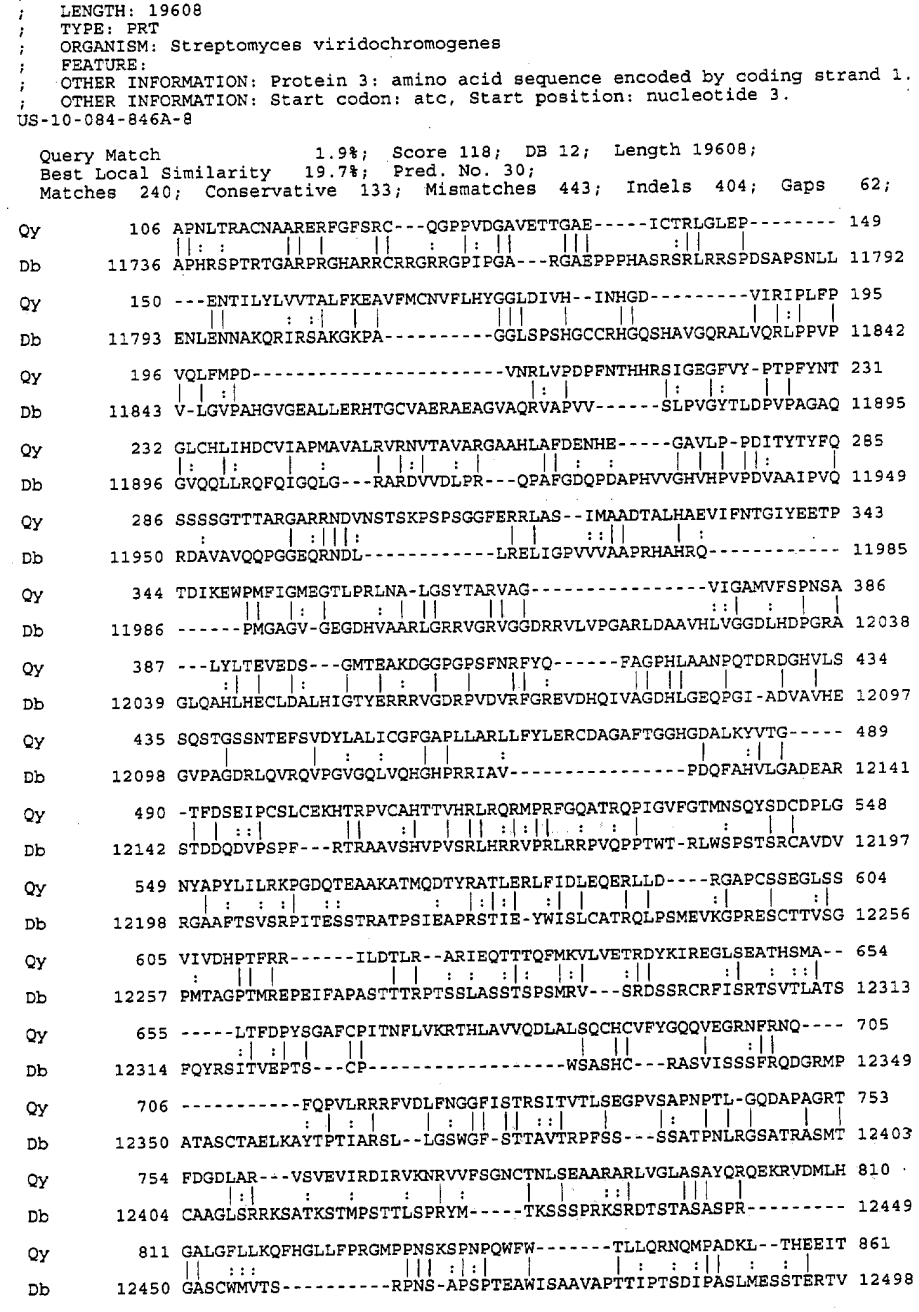
```
RESULT 13
US-09-991-630-13
; Sequence 13, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
;  APPLICANT: Einat, et al
   TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
   TITLE OF INVENTION:  AND USES THEREOF
   FILE REFERENCE: 540579-2007.3
   CURRENT APPLICATION NUMBER: US/09/991,630
   CURRENT FILING DATE:  2001-11-06
   PRIOR APPLICATION NUMBER: 09/905,129
   PRIOR FILING DATE: 2001-07-13
   PRIOR APPLICATION NUMBER: 09/802,318
   PRIOR FILING DATE: 2001-03-08
   PRIOR APPLICATION NUMBER: 09/729,485
   PRIOR FILING DATE: 2000-12-04
   NUMBER OF SEQ ID NOS: 28
   SOFTWARE: PatentIn version 3.0
 SEQ ID NO 13
    LENGTH: 2597
    TYPE: PRT
    ORGANISM: Rattus species
    FEATURE:
    NAME/KEY: misc_feature
    LOCATION: (1)..(2597)
    OTHER INFORMATION: 'x' can be any amino acid
S-09-991-630-13

 Query Match             1.9%;  Score 120.5;  DB 10;  Length 2597;
 Best Local Similarity   18.6%;  Pred. No. 0.46;
 Matches  145;  Conservative  113;  Mismatches  287;  Indels  233;  Gaps   40;

y       416 AGPHLAANPQTD---RDGHVLSSQSTGSSNTE-----------------FSVDYLALICG 455
            :|  |   |  |    :   ::: :  ||::|                 | :   ::|
b       274 SGAFLCTKPTIDPSLKSKSLVTQEDNGSASTSPQDFIEPPFGSLSLNMTXXSGNKADMVCS 333

y       456 FGAPLLARLLFYLERCDAGAFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV----CAHT 511
              |          |    |||  :   :  :  :| : :||:   | :||      ::
b       334 IQKP---------SRTSPTAFTEENDYIM--LNASFSTNLVCSVDYNHIQPVWQLLALYS 382

y       512 TVHRLRQRMP----------RFGQATRQPIGVFGTMNSQYSDCDPLGNYAPYLILRKPGD 561
              : :| |           |: |   :|  :| :: :      ||          ::|:
b       383 DSPLILERKPQLTETPSLSSRYKQVALRPEDIFTSIEADVR-ADPFWFQQEKIVLQLNRT 441

y       562 QTEAAKATMQDTYRATLERLFIDLEQERL-----------LDR--------GAPCSSEG 601
            |  :   :| :  | :     ::  |||             |:|                | :|
b       442 ATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNNPKLERTVLVGGTIALSCPGKG 501

y       602 -----LSSVIVDHPTFR--------RILDTLRARIE-QTTTQFMKVL---VETRD----- 639
                |  :: |      |          |||      ::| |      |    |  : | |
b       502 DPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQMADSFDAGLYHCISTNDADADV 561

y       640 --YKIR--EGLSEATHSMALTFDPYSGAF----CPITNFLVKRTHLAVVQDLALSQCHCV 691
              |:|    |   |:||   :     :|       | |             | | ::|    :
b       562 LTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTG----------VPDASIS---WI 608

y       692 FYGQQVEGRNFRNQFQPVLRRRFVDLFNGGFISTRSITVTLSEGPVS--APNPT------ 743
            | | :  |::          : | | :    :|    :|     | ||:
b       609 LPGNTVFSQPSRDR----------QILNNGTLRILQVT-PKDQGHYQCVAANPSGADFSS 657

y       744 ------------LGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCTNLSEAARAR 791
                      :  |  || :  |:     || : :   :|      | :    ||| :
b       658 FKVSVQKKGQRMVEHDREAGGSGLGE-PNSSVSLKQPASLK----LSASALTGSEAGKQ- 711

y       792 LVGLASAYQRQEKRVDMLHGALG-FLLKQFHGLLFPRGMPPNSKSPNPQWFWTLLQ---R 847
              |   |: |  |::|    |  |::|      | :| ::: :|| :  ||:   :
b       712 ----VSGVHRKNKHRDLIHRRRGDSTLRRFRE--HRRQLPLSARRIDPQRWAALLEKAKK 765

y       848 NQMPADKLTHEEITT-------IAAVKRFTEEYAAINFINLPPTCIGELAQFYMANLILK 900
            | :|      :| ||       :  |:   ||  |    |  ||
b       766 NSVP----KKQENTTVKPVPLAVPLVELTDEEKDASGMI--PP---------------- 802

Qy      901 YCDHSQYLINTLTSIITGARRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELW 960
              |    :: |  | : |     |  | |     : | :: |      |: :|:||:
Db      803 --DEEFMVLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPQTLQ-SEHLPDFK 859

Qy      961 TTAFTSTHLVRAAM-----------NQRPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
              : |:   |  :|            || |:::   |:::   :|     |||  |
Db      860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIIFP-SVAEIRDSA------QAGRASSQS 912

Qy     1007 ---LNGGKNVCPLFTFDRTRRFIIACPRGGFI-----------CPVTG---PSSGN 1045
              : ||     : |:  |  :     :   :                |:||   ||| :
Db      913 AHPVTGGN----MATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGVSRPSSSD 966


RESULT 14
US-10-329-079-11
; Sequence 11, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
;  APPLICANT: FARNET, Chris
;  APPLICANT:  ZAZOPOULOS, Emmanuel
;  APPLICANT:  STAFFA, Alfredo
;  TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
;  FILE REFERENCE: 3002-11US
;  CURRENT APPLICATION NUMBER: US/10/329,079
;  CURRENT FILING DATE:  2002-12-24
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
;   LENGTH: 5245
;   TYPE: PRT
;   ORGANISM: Streptomyces fradiae
US-10-329-079-11

  Query Match             1.9%;  Score 120;  DB 12;  Length 5245;
  Best Local Similarity   20.2%;  Pred. No. 1.8;
  Matches  262;  Conservative  138;  Mismatches  434;  Indels  464;  Gaps   68;

Qy      102 PSSAAPNLTRACNAARERFGFSRCQGPP----VDGAVETTGAEICTRL----------GL 147
            |::| |          | :|      |  |  | |  | |  |  |:          |
Db      464 PAAAEPAADEGLEAVCDTFARQAAATPEAPAVVGGPVALTFAEADARVSRLARLLISRGA 523

Qy      148 EPENTILYLVVTALFKEAVFMCNVFLHYGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLV 207
              ||      |  | :|::  |          : :      : :||
Db      524 GPE----VRVAVCLDRNALWPTTV---------LAVLRSGAVHVPL------------- 556

Qy      208 PDPFNTHHRSIGEGFVYPTPFYNTGLCHLIHDCVIAPMAVALRVRNVTAVARGAAHLAF- 266
            || : | |                | :  | :||:|          ||| || : : .
Db      557 -DPRSPHER---------------LAAVERD--VAPLLVLAERATEAAVADLAAPVLVL 597

Qy      267 -DENHEGAV----------------LPPDITYTYFQSSSSG-----TTTARG------A 297
             | : | |:                ||     |  | |:|     |    ||      |
Db      598 DDPSTEAAIDALDPGPVTDADRTAPLLPGHAAYVIHTSGSTGRPKGVTVDHRGLSRLLQA 657

Qy      298 RRNDVNSTSKPSPSG-GFERRLASI----------MAADTALHA-------------- 330
             |    | :||  | |    ::|             |  |  ||
Db      658 HRRVTFSRIRPSAGGPGRAAHVSSFSFDASWDPLLAMVAGHELHMIDEDLRFDPPGVVAY 717

Qy      331 ---------------EVIFNTGIYEE---TPT------DIKEWPMFIGMEGTLPRLNAL 365
                          : : |: |||       :  :  |:  :  :: :  ||: |:
Db      718 FRDRRIDYVDLTPTYFRSLLDAGLLEEGFPCPSLVALGGEAMDGELWERLRAAAPRVTAM 777

Qy      366 GSY----TARVA----------GVIGAMVFSPNSALYLTEVEDSGMTEAKDGGPGPSFNR 411
              :|    ||  |          | || |  |  |  | |:|:     |>| |:
Db      778 NTYGPTETAVDAVVTVLGDLPPGTIGRPV--PRWRAY---VLDAGLRPVPPGVLGELY-- 830

Qy      412 FYQFAGPHLAANPQTDRDGHVLSSQSTGSSNTEFSVDYLALICGFGAP--------LLAR 463
             ||| :|          | |:::          : | || |          |||
Db      831 ---LAGPGVARGYLGQ---HALTAER-----------FVACPFGKPGERMYRTGDLAR 871

Qy      464 L-----LFYLERCD-------------------------AGAF-----TGGHGDALKYV 487
                  | |: | |                         |  |     | |    : ||
Db      872 WLPDGHLVYVGRGDEQVKIRGFRIEPGEVEAALRELEGVAAAAVTVREDTPGTRRLVGYV 931
```



RESULT 15
US-10-084-846A-8
; Sequence 8, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2

; SEQ ID NO 8
; LENGTH: 19608
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-8

Query Match 1.9%; Score 118; DB 12; Length 19608;
Best Local Similarity 19.7%; Pred. No. 30;
Matches 240; Conservative 133; Mismatches 443; Indels 404; Gaps 62;



```
  862 TIAAVKRFTEEYAAI--NFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSIITGA 919
      :|        |    |  :    |||                    : | :: :|| |
12499 LLATGTSCLAEVCVIGRSLEPTPPT-----------------NMSAFISHTLRSA---- 12536

  920 RRPRDPSSVLHWIRKDVTSAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAMNQRPM 979
          ||  :  : :::  |:::  | ::     :  |          |          |
12537 -----PSDQV--VSRELQMASNVSAQQRSYWRVFQVFPRFPGRLARPFHPAVPCHELRSQ 12589

  980 VV-------LGISISKYHGAAGNNRVFQAGNWSGLNGGKNVCPLFTFDRTRRFIIACPRG 1032
       |       :| |: :   :  :       |::  :  | :   .     | |||
12590 CVQDAWPSAVGCSVRR---SVRSTMSLMVGSFQKMLAGTS--------RPRRF------- 12631

 1033 GFICPVTGPSSGNRETTLSDQVRGIIVSGGAMVQLAIYATVVRAVGARAQHMAFDDWLSL 1092
             ||  | ||:  :      ||            |||||     |        |
12632 ---------SSRRRTTTVITESSPCSTSG---------RRTCRAVGASRISCAI---CSR 12670

 1093 TDDEFLARDLEELHDQIIQTLETPWTVEGALEAVKILDEKTTAGDGETPTNLAFNFDSCE 1152
      |                                    |::::||
12671 TSSS--------------------------------TRERSSSGD--------------R 12684

 1153 PSHDTTSNVLNISGSNISGS 1172
      |::||||    :  ||::||:
12685 PANDTTSASASPPGSSVSGA 12704
```

Search completed: January 30, 2004, 13:20:11
Job time : 49 secs

i903590KP1

ion date: 03-03-2004
king Officer: HABDULEKADER - HUDA ABDULEKADER
n: OIPEScanning
sier: 10190359

al Date: 09-16-2002

o.	Doccode	Number of pages
	NPL	12
	NPL	4
3	NPL	5
4	NPL	8
5	NPL	9
6	NPL	4
7	NPL	5
8	NPL	2
9	NPL	13
10	NPL	6
11	NPL	14
12	NPL	6
13	NPL	8
14	NPL	40
15	NPL	6
16	NPL	7
17	NPL	8
18	NPL	4
19	NPL	3
20	NPL	4
21	NPL	30
22	NPL	8
23	NPL	4
24	NPL	8
25	NPL	17
26	NPL	14
27	NPL	5
28	NPL	9
29	NPL	12
30	NPL	7
31	NPL	10